AWARD NUMBER: W81XWH-13-1-0319

TITLE: A Novel Approach to Assay DNA Methylation in Prostate Cancer

PRINCIPAL INVESTIGATOR: Jindan YU

CONTRACTING ORGANIZATION: Northwestern University
Chicago, IL 60611

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2015	Annual	30 Sep 2014 - 29 Sep 2015

4. TITLE AND SUBTITLE

A Novel Approach to Assay DNA Methylation in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-1-0319

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jindan Yu

E-Mail: jindan-yu@northwestern.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Northwestern University
303 E. Superior St.
Chicago, IL 60611

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Previous studies of DNA methylation at 5-position of cytosine (5mC) have led to the discovery of useful methylation biomarkers for prostate cancer diagnosis and prognosis, some of which are being developed into clinical tests. However, several seminal studies have recently reported that DNA methylation (5mC) can be de-methylated by the TET proteins resulting in 5-hydroxylmethylation (5hmC), which plays functional roles distinct from 5mC but yet is indistinguishable from 5mC by a majority of existing methylation assays. Developing enabling assays that measure 5mC and 5hmC specifically might significantly improve the performance of methylation biomarkers. During this reporting period, we have performed genome-wide mapping of 5mC and 5hmC in 33 samples, including 11 primary benign prostate tissue, 11 localized prostate cancer tissue and 6 castration-resistant prostate cancer tissue. We are working on to establish a bioinformatics system to handle and analyze the large amount of sequencing data. Further, in an effort to establish a system to manipulate DNA methylation, we found that FOXA1, a pioneer factor of androgen receptor, directly induces TET1 expression and interacts with TET1 protein, leading to lineage-specific enhancer activation. Several manuscripts have been either published or under submission.

15. SUBJECT TERMS

DNA methylation 5mC, DNA hydroxymethylation 5hmC, biomarker, 5hmCSL-PCR, 5mCSL-PCR, BS-seq, TAB-seq, cancer-specific methylation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	69	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. Introduction…….4

2. Keywords…….4

3. Accomplishments ……... 4

4. Impact……...……12

5. Changes/Problems...…..……12

6. Products……..……..…..…….13

7. Participants & Other Collaborating Organizations ……13

8. Special Reporting Requirements……15

9. Appendices……16

1. **INTRODUCTION:**

Previous studies of DNA methylation at 5-position of cytosine (5mC) have led to the discovery of useful methylation biomarkers for prostate cancer diagnosis and prognosis, some of which are being developed into clinical tests. However, several seminal studies have recently reported that DNA methylation 5mC can be de-methylated by the TET proteins resulting in 5-hydroxylmethylation (5hmC), which plays distinct functional roles of 5mC but yet is indistinguishable from 5mC by a majority of current assays. Developing enabling assays that measure 5mC and 5hmC specifically might significantly improve the performance of methylation biomarkers. Our goal is thus to develop and apply 5mC- or 5hmC-specific assays to study prostate cancer (PCa) methylome. Our hypothesis is that 5hmC and 5mC play distinct and important roles in PCa and cancer-specific 5hmC and 5hmC signatures may be useful PCa biomarkers. To test this hypothesis, we propose two Specific Aims:

(1) Develop 5mC- and 5hmC-specific assays for DNA methylation analysis in PCa.

(2) Obtain genome-wide maps of 5mC and 5hmC distribution in prostate cancer.

2. **KEYWORDS:** DNA methylation 5mC, DNA hydroxymethylation 5hmC, biomarker, 5hmCSL-PCR, 5mCSL-PCR, BS-seq, TAB-seq, cancer-specific methylation

3. **ACCOMPLISHMENTS:**

- **What were the major goals of the project?** The Table below listed the goals/target dates of the project as in the approved SOW. A new column was added on the right showing actual completion dates and the percentage of completion.

Specific Aim 1. To develop 5mC- and 5hmC-specific assays for DNA methylation analysis in PCa	**Timeline**	**Site 1**	**Site 2**	**% of completion**	**Actual complete dates**
Major Task 1 : Develop 5hmCSL-PCR assay	Months				
Subtask 1: optimize 5hmCSL-PCR in a panel of prostate cell lines using previously known methylated regions	1-6	Dr. Yu	Dr. He	100%	9/1/2015
Subtask 2: Compare 5hmCSL results with hMeDIP.	6-12	Dr. Yu		80%	
Subtask 3: Use TAB-pyro to validate 5hmCSL results.	6-12	Dr. Yu		30%	
Local IRB non-human subject Approval	1-3			100%	3/22/2014
Milestone Achieved: HRPO Approval	6			100%	3/22/2014
Major Task 2: Develop 5mCSL-PCR assay					
Subtask 1: optimize 5mCSL-PCR in a panel of prostate cell lines using previously known	3-9	Dr. Yu	Dr. He	80%	

methylated regions					
Subtask 2: Compare 5mCSL results with MeDIP.	6-12	Dr. Yu		100%	6/31/2014
Subtask 3: Use BS-pyro and TAB-pyro to validate 5mCSL results.	9-18	Dr. Yu		0%	
Major Task 3: Assess methylation biomarkers in prostate cancer specimens.					
Subtask 1: analyze 5mC and 5hmC level of selected genes using DNA from 90 prostate cancer tissues.	12-18	Dr. Yu		50%	
Specific Aim 2: Obtain genome-wide map of 5mC and 5hmC distribution in prostate cancer					
Major Task 4: Genome-wide 5(h)mCSL-Seq					
Subtask 1: perform 5(h)mCSL-Seq in 3 cell lines.	9-18	Dr. Yu	Dr. He	70%	
Subtask 2: identify cancer-specific 5(h)mC loci.	12-30	Dr. Yu		50%	
Major Task 5: BS-Seq and TAB-Seq					
Subtask 1: BS-Seq and TAB-Seq in 3 cell lines	24-30	Dr. Yu	Dr. He	50%	
Subtask 2: identify 5(h)mC-rich loci and cross-validation using different assays.	30-36	Dr. Yu		30%	
Subtask 3: analyze cancer-specific 5(h)mC biomarker using DNA from 90 prostate cancer tissues.	30-36	Dr. Yu		0	

- **What was accomplished under these goals?**

1) Major activities in this reporting period include:

A. Perform genome-wide 5mCSL-seq to map 5mC in 33 prostate cancer samples, including 5 cell lines, 11 benign prostate tissues, 11 localized prostate cancer, and 6 metastatic prostate cancer tissues.

B. Perform genome-wide 5hmCSL-seq to map 5hmC in 33 prostate cancer samples, including 5 cell lines, 11 benign prostate tissues, 11 localized prostate cancer, and 6 metastatic prostate cancer tissues.

C. Determine how TET1 occupancy and 5hmC modification are associated with enhancer activities.

D. Investigate how FOXA1 regulates TET1 expression and function.

E. Examine how TET1 facilitates FOXA1-mediated enhancer activation through DNA demethylation.

2) Specific objectives: (A) Optimize existing and novel methylation assays; (B) obtain genome-wide maps of 5mC and 5hmC.

3) Significant results or key outcomes:

Major Task 1: Develop 5hmCSL-PCR assay.

Progress: The 5hmCSL-PCR assay has been successfully developed (subtask 1). To compare the performance of 5hmCSL and hMeDIP (subtask 2), we have performed 5hmCSL-qPCR and hMeDIP-qPCR on a set of genes and found that the former is much more sensitive and accurate in capturing 5hmC (**Figure 1** below). The 5hmCSL-qPCR assay, as expected, showed reduced 5hmC levels at target

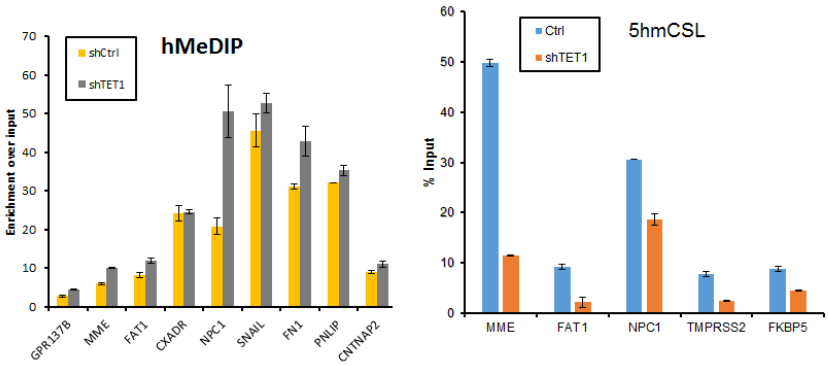


Figure 1: 5hmCSL-PCR is more accurate in determining 5hmC level than hMeDIP. LNCaP cells with control and TET1 knockdown were subjected to hMeDIP and 5hmCSl followed by qPCR analysis.

methylated regions following knockdown of TET1, the DNA demethylase that converts 5mC to 5hmC. This may be due to the fact that 5hmC only exists on a small percentage (~2.5%) of Cs and may be insufficient to enrich by antibody-based approach (the hMeDIP assay).

Major Task 2: Develop 5mCSL-PCR assay

Progress: we have developed the 5mCSL-PCR assay (subtask 1). In order to compare the effectiveness of 5mCSL assay in relative to MeDIP (subtask 2), we performed next-generation sequencing of the enriched DNA. We found that with comparable amount of total sequencing reads, 5mCSL-seq leads to very few and indistinct peaks when compared to MeDIP-seq which render sharp and clear peaks of enrichment (**Figure 2**). This may be due to the fact that 5mC is quite abundant and the level of modified 5mC is sufficient for antibody-based enrichment. On the other hand, 5mCSL requires blockade of original 5hmC and conversion of 5mC to new 5hmC for subsequent chemical-linked pull down, a process that may result in substantial loss of the material.

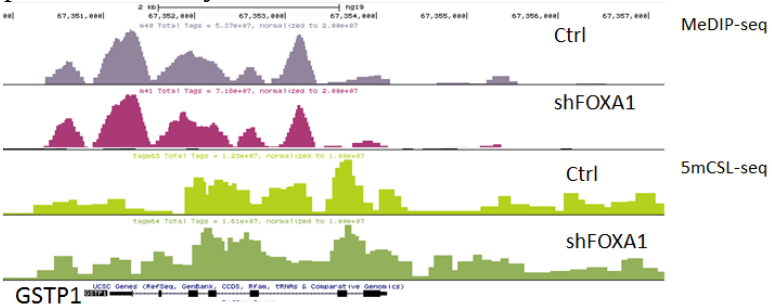


Figure 2: MeDIP-seq outperformed 5mCSL-seq. LNCaP cells with control and FOXA1 knockdown were subjected to MeDIP and 5mCSL pull down followed by NGS analysis. The enriched peaks around the GSTP1 gene are shown.

Consequently, it is no longer needed to compare 5mCSL based assay with pyrosequencing (subtask 3). We have thus decided to use MeDIP-seq in the later part of the project for genome-wide analysis of 5mC, and 5hmCSL-seq for genome-wide mapping of 5hmC.

Major Task 3: Assess methylation biomarkers in prostate cancer specimens.

Progress: we initially proposed to analyze a set of known methylated genes/loci in 90 benign adjacent and prostate cancer tissues. However, as we do not have a well-defined set of 5hmC-enriched genes and the cost for NGS has dropped quite dramatically, we have decided to perform global analysis of 5mC and 5hmC in primary samples to identify cancer-specific 5mC and 5hmC methylation. As of today, we have successfully carry out genome-wide mapping of 5mC and 5hmC in primary tissue specimens as listed below (**Table 1**). The data are currently being collected and analyzed. Some of the samples need additional sequencing to reach sufficient coverage, which is currently ongoing. After we complete sequencing of these specimens, we will build bioinformatics pipelines to determine cancer-specific 5mC and 5hmC regions. If necessary, additional specimens will be subjected to genome-wide analysis. Cancer-specific methylation biomarkers will be determined.

Table 1: Next-generation sequencing analysis of DNA methylation (5mC and 5hmC) in primary specimens.

Sample ID	Cell Line	Treatment	Antibody	Barcode	Reads
9068-21	normal	tissue	5hmC-seq	701-502	5,464,789
31443	normal	tissue	5hmC-seq	702-503	3,379,152
31155	normal	tissue	5hmC-seq	703-504	4,328,516
9050-24	normal	tissue	5hmC-seq	704-505	6,007,179
9102-21	normal	tissue	5hmC-seq	705-506	3,308,939
30833	normal	tissue	5hmC-seq	706-507	4,526,880
31677	normal	tissue	5hmC-seq	707-508	3,682,219
38597	normal	tissue	5hmC-seq	708-517	4,231,648
49246	normal	tissue	5hmC-seq	709-502	5,304,758
38545	normal	tissue	5hmC-seq	710-503	2,925,850
Input1	normal	tissue	5hmC-seq	703-503	2,118,178
9068-23	G7/91-100%	tissue	5hmC-seq	711-504	4,406,930
31154	G9/61-70%	tissue	5hmC-seq	712-505	9,075,967
31444	G9/71-80%	tissue	5hmC-seq	701-506	12,399,696
9050-23	G7/51-60%	tissue	5hmC-seq	702-507	6,456,052
9102-24	G9/20-30%	tissue	5hmC-seq	703-508	11,646,542
30831	G9/41-50%	tissue	5hmC-seq	704-517	8,068,573
31678	G9/61-70%	tissue	5hmC-seq	705-502	6,236,421

38600	G9/71-80%	tissue	5hmC-seq	706-503	5,370,120
49245	G9/61-70%	tissue	5hmC-seq	707-504	3,622,852
38547	G9/21-30%	tissue	5hmC-seq	708-505	6,004,212
Input 2	cancer	tissue	5hmC-seq	704-504	3,599,877
27	LNCaP	control	5hmC-seq	701-503	7,591,719
28	LNCaP	shTET1	5hmC-seq	702-504	5,044,257
DMSO	LNCaP	DMSO	5hmC-seq	703-505	5,365,282
MDV-R	LNCaP	MDV-R	5hmC-seq	704-506	2,772,933
Input 4	LNCaP	none	5hmC-seq	706-506	3,613,131
31-27	CRPC	tissue	5hmC-seq	709-506	6,226,398
37-54	CRPC	tissue	5hmC-seq	710-507	3,464,594
30-28	CRPC	tissue	5hmC-seq	711-508	6,442,382
28-34	CRPC	tissue	5hmC-seq	712-517	3,287,285
33-89	CRPC	tissue	5hmC-seq	702-502	2,385,396
Input3	CRPC	tissue	5hmC-seq	705-505	3,153,956
38597	Normal	tissue	MeDIP-seq	GGCTAC	25,813,531
30-28	CRPC	tissue	MeDIP-seq	ATCACG	31,184,661
33-89	CRPC	tissue	MeDIP-seq	TTAGGC	23,055,143
9050-23	G7/51-60%	tissue	MeDIP-seq	CAGATC	44,933,849
9102-24	G9/20-30%	tissue	MeDIP-seq	ACTTGA	72,867,001
30831	G9/41-50%	tissue	MeDIP-seq	GATCAG	29,065,739
27	LNCaP	control	MeDIP-seq	ACAGTG	15,886,470
28	LNCaP	shTET1	MeDIP-seq	GCCAAT	20,621,914
30833	normal	tissue	MeDIP-seq	706-507	18,033,668
31677	normal	tissue	MeDIP-seq	707-508	21,259,353
38597	normal	tissue	MeDIP-seq	708-517	23,884,469
49246	normal	tissue	MeDIP-seq	709-502	26,091,306
38545	normal	tissue	MeDIP-seq	710-503	19,426,891
Input1	normal	tissue	MeDIP-seq	703-503	21,826,000
30831	G9/41-50%	tissue	MeDIP-seq	704-517	21,644,173
31678	G9/61-70%	tissue	MeDIP-seq	705-502	24,138,360
38600	G9/71-80%	tissue	MeDIP-seq	706-503	21,329,426
49245	G9/61-70%	tissue	MeDIP-seq	707-504	24,840,686
38547	G9/21-30%	tissue	MeDIP-seq	708-505	23,056,205
31-27	CRPC	tissue	MeDIP-seq	709-506	30,419,917
37-54	CRPC	tissue	MeDIP-seq	710-507	21,981,475
30-28	CRPC	tissue	MeDIP-seq	711-508	22,645,996
28-34	CRPC	tissue	MeDIP-seq	712-517	23,878,434
33-89	CRPC	tissue	MeDIP-seq	702-502	23,129,750
30833	normal tissue	tissue	MeDIP-seq	M198	55,178,693
37-54	CRPC tissue	tissue	MeDIP-seq	M201	39,691,324
28-34	CRPC tissue	tissue	MeDIP-seq	M203	42,248,520
9068-23	PC tissue	tissue	MeDIP-seq	M205	41,202,546

31154	PC tissue	tissue	MeDIP-seq	M206	32,865,094
31444	PC tissue	tissue	MeDIP-seq	M207	42,734,690

Methods: *Methylated DNA Immunoprecipitation (MeDIP):* Briefly, total genomic DNA was extracted using QIAamp DNA Mini Kit (Qiagen), and sonicated to obtain fragments between 300-1000bp. Dynabeads M-280 Sheep anti-Mouse IgG (Invitrogen) were incubated with an anti-5-methylcytidine antibody (BI-MECY_0100, Eurogentec, Fremont, CA, USA) overnight at 4^{o}C. The following day, 4ug of sheared DNA was denatured by boiling at 95 ^{o}C for 10min followed by rapid cooling on ice, and subsequently added to the beads/antibody complex. On day 3, the beads were washed 3 times with PBS+0.05% Triton X-100 and eluted from beads by incubation at 65^{o}C for 5min in 150ul elution buffer (TE+1% SDS). Elution was repeated for a total of two times. Total eluates were treated with proteinase K and incubated at 50^{o}C for 2hr. QIAquick PCR purification kit (Qiagen) was used to purify the eluted DNA, and lastly qPCR was used to determine the enrichment of target genomic regions using gene-specific primers (listed in Supplemental Information). Enrichment of target loci was normalized to input DNA and reported as % input ±SEM.

5mC and 5hmC Chemical Labeling (TAmC and hMe-Seal): 5mC labeling experiments were done as previously published (Zhang, Szulwach et al. 2013), and 5hmC labeling experiments were performed as previously described (Song, Clark et al. 2012). Briefly, genomic DNA was fragmented to an average of 400bp and was incubated with 50mM HEPES buffer (pH 7.9), 25mM MgCl2, 100mM UDP-6-N3-Glc, and 2mM βGT for 1hr at 37°C. The labeled DNA was purified by the QIAquick Nucleotide Removal kit (Qiagen) and eluted in H2O. The click chemistry was performed with the addition of 150mM of disulfide-biotin, and the mixture was incubated for 2hr at 37°C. The labeled DNA fragments were then purified by the QIAquick Nucleotide Removal kit (Qiagen) and enriched by Dynabeads Streptavidin C1 (Invitrogen), and subsequently released by DTT treatment. The enriched DNA fragments were first purified by Micro Bio-Spin 6 spin columns (Bio-Rad) followed by MinElute PCR Purification Kit (QIAGEN).

Major Task 4: Genome-wide 5(h)mCSL-Seq in 3 cell lines.

This has been completed as shown in the table below.

Cell line	Condition	Method	Barcode	Name	#seq reads
PrEC LHS	Normal	5mC pull down	CGATGT	A1	27,462,926
LNCaP	Normal	5mC pull down	TGACCA	A2	28,687,039
PC-3M	Normal	5mC pull down	ACAGTG	A3	24,459,274
PrEC LHS	Normal	5hmC pull down	GCCAAT	A4	29,953,414
LNCaP	Normal	5hmC pull down	CAGATC	A5	25,075,174
PC-3M	Normal	5hmC pull down	CTTGTA	A6	30,563,514

In addition, during the analysis of this data, we found that 5mC is depleted whereas 5hmC enriched at FOXA1 binding sites in the AR+, FOXA1+ LNCaP cells, but not in the PrEC and PC-3M cells (**Figure 3** below). We thus hypothesized that **FOXA1 may be a critical regulator of TET1 and/or DNA demethylation**. To investigate this, we have examined how FOXA1 regulates TET1 expression and enhancer function. The key findings are summarized below and the manuscript reporting the results is currently under revision (**Appendix 2**).

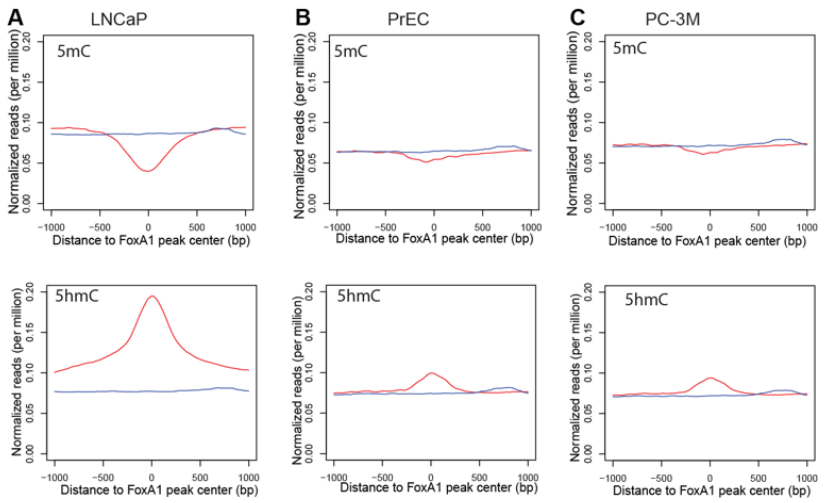
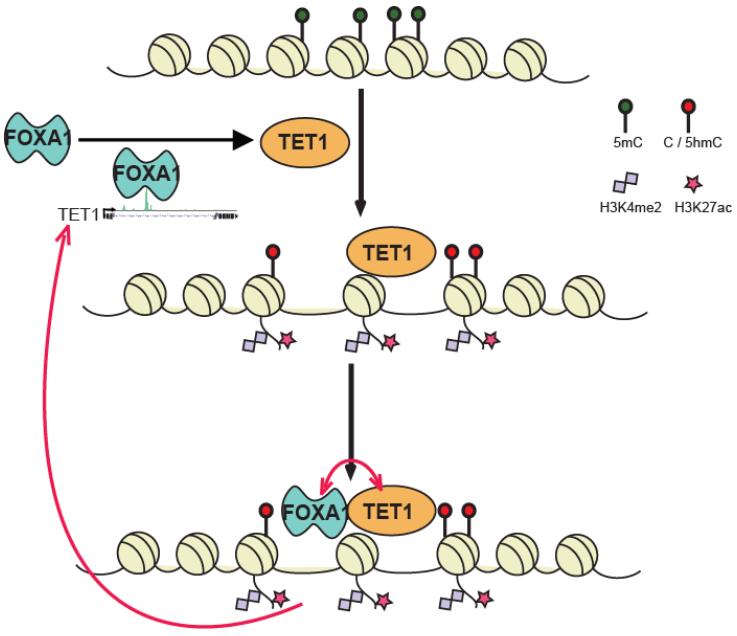


Figure 3: Intensity plots showing 5mC and 5hmC enrichment around FOXA1 binding sites (±1 kb) in LNCaP (A), PrEC (B), and PC-3M cells (C). MeDIP and 5hmCSL were performed using genomic DNA extracted from LNCaP, PrEC and PC-3M cell lines. Enriched DNA was made into libraries and subjected to deep sequencing. The read intensities of were evaluated relative to FOXA1 binding sites in LNCaP cells.

<u>Investigate how FOXA1 regulates lineage-specific enhancers through modulating TET1.</u>

Results:

1. FOXA1 contributes to enhancer activation through epigenetic modifications (**Figure 1 of Appendix 2**).
2. FOXA1 induces TET1 gene expression (**Figure 2 of Appendix 2**).
3. TET1 is a direct transcriptional target of FOXA1 (**Figure 3 of Appendix 2**).
4. FOXA1 and TET1 proteins physically interact (**Figure 4 of Appendix 2**).
5. TET1 mediates active epigenetic modification at FOXA1-bound enhancers (**Figure 5 of Appendix 2**).
6. TET1 is required for FOXA1 recruitment to lineage-specific enhancers (**Figure 6 of Appendix 2**).



Conclusion (Figure on the right): FOXA1 protein occupies at an intragenic enhancer of the TET1 gene to directly induce its expression. Through direct interaction with FOXA1 protein, TET1 modulates DNA demethylation and subsequently H3K4 methylation and H3K27 acetylation at FOXA1-target enhancers, which in turn facilitates FOXA1 recruitment. Thus, FOXA1 and TET1 form a positive feedback loop

in lineage-specific enhancer activation. FOXA1 is not only a reader, but also a writer of epigenetic signatures at lineage-specific enhancers.

Major Task 5: BS-Seq and TAB-Seq in RWPE, LNCaP and PC-3M cells.

Progress: We have performed this in one cell line (LNcaP) with control and knockdown of FOXA1 (which is to deregulate 5mC level) as shown below. In order to obtain base-level methylation information, we may need to do deep sequencing to get 1 billion reads per sample. In addition, we are developing computational programs for the analysis and visualization of the methylation data. Once this is achieved, we will determine cancer-specific 5mC/5hmC and compare the base-level data to enrichment-based methylation data obtained through 5hmCSL-seq and MeDIP-seq.

Cell line	Condition	Method	barcode	Sample	# seq reads
LNCaP	shGIPZ BS	BS-seq	GGTAGC	MK126	235,589,814
LNCaP	shFOXA1 BS	BS-seq	ATGAGC	MK127	288,264,046
LNCaP	shGIPZ TAB	TAB-seq	TATAAT	MK128	306,755,767
LNCaP	shFOXA1 TAB	TAB-seq	TCGAAG	MK129	273,079,194

Write and publish a research article on how an enhancer-regulated (through FoxA1 and potentially TET1) lncRNA HOTAIR regulates estrogen receptor signaling in breast cancer (**Appendix 1**). This grant support was thus acknowledged in the published paper.

- **What opportunities for training and professional development has the project provided?**

"Training" activities:

At the Northwestern University, bioinformatician Jonathan Zhao spent 3.6 calendar months on the analysis of the next-generation sequencing data, postdoctoral fellow Dr. Ali Zhang and Bing Song have spent 2.4 and 6 calendar months working on the project, respectively. In addition, a graduate student Angela Yang, supported by a T32 training grant, has worked on this project for 10 calendar months. They have received one-on-one training by discussion and meetings with the PIs and also by collaborating with the He's laboratory. They have gained extensive training in the study of DNA methylation and in various methylation assays.

At the University of Chicago, Mr. Xingyu Lu and Ms. Miao Yu (HHMI international pre-doctoral fellow) have been involved in the research. They have gained significant knowledge on prostate cancer research through interactions with the Yu laboratory. A new postdoc, Dr. Lulu Hu, has taken over the project and started to learn prostate cancer research.

"Professional development" activities:

The PI Dr. Yu has given lectures in various universities, including University of Michigan, Methodist Hospital, Mayo Clinic, and Harvard University on related topics.

Professor Chuan He has been given lectures on 5hmC in various scientific meetings and visits of other schools.

- **How were the results disseminated to communities of interest?**

Nothing to Report

- **What do you plan to do during the next reporting period to accomplish the goals?**

During the next reporting period, we plan to do the following:

(1) Optimize and validate various methylation assays to confirm genome-wide findings.

(2) Deep sequencing of 5mC and 5hmC to reach desired coverage and in more primary specimens

(3) Bioinformatic analysis of all high-throughput sequencing data.

(4) Optimize 5mCSL and 5hmCSL using small amount of DNA.

(5) Identify cancer-specific 5mC and 5hmC methylation biomarkers.

4. **IMPACT:**

- **What was the impact on the development of the principal discipline(s) of the project?**

Nothing to Report

- **What was the impact on other disciplines?**

Nothing to Report

- **What was the impact on technology transfer?**

Nothing to Report

- **What was the impact on society beyond science and technology?**

Nothing to Report.

5. **CHANGES/PROBLEMS:**

- **Changes in approach and reasons for change**

Nothing to Report

- **Actual or anticipated problems or delays and actions or plans to resolve them**

Nothing to Report

- **Changes that had a significant impact on expenditures**

Nothing to Report

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to Report

6. **PRODUCTS:**

- **Publications, conference papers, and presentations**

- **Journal publications.**

1. Xue XY, Yang YA, Zhang A, Fong KW, Kim J, Song B, Li S, Zhao JC, Yu J. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2015 Sep 14. PMID: 26364613 [Epub ahead of print]
2. Yang YA, Zhao JC, Fong KW, Kim J, Li S, Song C, Song B, Zheng B, He C, **Yu J**. *FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function.* **Under Review**.

- **Books or other non-periodical, one-time publications.**

Nothing to Report

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

Nothing to report

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

Nothing to report

7. **PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

- **What individuals have worked on the project?**

Name:	*Jindan Yu*

Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	*1071255*
Nearest person month worked:	*1.2*
Contribution to Project:	*Dr. Yu supervised all the work related to this project, designed the experiments, facilitated collaboration between researchers, write reports and manuscripts.*
Funding Support:	

▪

Name:	*Angela Yang*
Project Role:	*Graduate Student*
Researcher Identifier (e.g. ORCID ID):	*2654097*
Nearest person month worked:	*10*
Contribution to Project:	*Angela performed the work related to MeDIP and hMeDIP pull down assays, performed cell culture, DNA isolation, and methylation assays. She also performed analysis of the next-generation sequencing data by checking the results in genome-browser.*
Funding Support:	T32 Carcinogenesis training grant

Name:	*Ali Zhang*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	*1079458*
Nearest person month worked:	*2.4*
Contribution to Project:	*Dr. Zhang has performed work related to molecular cloning, cell culture and PCR assays.*
Funding Support:	

Name:	*Bing Song*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier	*1074460*

(e.g. ORCID ID):	
Nearest person month worked:	*6*
Contribution to Project:	*Dr. Bing has prepared all sequencing libraries for next-generation sequencing using the Illumina sample prep protocol and performed work in arranging and receiving the sequences.*
Funding Support:	

Name:	*Jonathan Zhao*
Project Role:	*Bioinformatician*
Researcher Identifier (e.g. ORCID ID):	*1082016*
Nearest person month worked:	*3.6*
Contribution to Project:	*Dr. Zhao is responsible for bioinformatics analysis of all next-generation sequencing data.*
Funding Support:	

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Nothing to Report

- **What other organizations were involved as partners?**

Organization Name: ***University of Chicago***

Location of Organization: *(if foreign location list country): Chicago, IL*

Partner's contribution to the project *(identify one or more): They performed the 5mCSL and 5hmCSL pull down involved in all experiments and arranged for part of the next-generation sequencing.*

Collaboration *(e.g., partner's staff work with project staff on the project)*; Mr. Xingyu Lu and Ms. Miao Yu at the University of Chicago participated in the 5mCSL and 5hmCSL pull down assays in prostate cancer DNAs that the Northwestern lab isolated and provided.

Personnel exchanges *(e.g., project staff and/or partner's staff use each other's facilities, work at each other's site)*; and

8. **SPECIAL REPORTING REQUIREMENTS**

COLLABORATIVE AWARDS: *NA.*

QUAD CHARTS: *NA.*

9. **APPENDICES:**

Appendix 1: Xue XY, Yang YA, Zhang A, Fong KW, Kim J, Song B, Li S, Zhao JC, Yu J. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2015 Sep 14. PMID: 26364613 [Epub ahead of print]

Appendix 2: Yang YA, Zhao JC, Fong KW, Kim J, Li S, Song C, Song B, Zheng B, He C, **Yu J**. *FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function.* **Under Review**.

Oncogene (2015), 1–10
© 2015 Macmillan Publishers Limited All rights reserved 0950-9232/15
www.nature.com/onc

ORIGINAL ARTICLE

LncRNA *HOTAIR* enhances ER signaling and confers tamoxifen resistance in breast cancer

X Xue[1,2,4], YA Yang[2,4], A Zhang[2], K-W Fong[2], J Kim[2], B Song[2], S Li[2], JC Zhao[2] and J Yu[2,3]

Tamoxifen, an estrogen receptor (ER) antagonist, is the mainstay treatment of breast cancer and the development of resistance represents a major obstacle for a cure. Although long non-coding RNAs such as HOTAIR have been implicated in breast tumorigenesis, their roles in chemotherapy resistance remain largely unknown. In this study, we report that *HOTAIR* (HOX antisense intergenic RNA) is upregulated in tamoxifen-resistant breast cancer tissues compared to their primary counterparts. Mechanistically, HOTAIR is a direct target of ER-mediated transcriptional repression and is thus restored upon the blockade of ER signaling, either by hormone deprivation or by tamoxifen treatment. Interestingly, this elevated *HOTAIR* increases ER protein level and thus enhances ER occupancy on the chromatin and potentiates its downstream gene regulation. *HOTAIR* overexpression is sufficient to activate the ER transcriptional program even under hormone-deprived conditions. Functionally, we found that *HOTAIR* overexpression increases breast cancer cell proliferation, whereas its depletion significantly impairs cell survival and abolishes tamoxifen-resistant cell growth. In conclusion, the long non-coding RNA *HOTAIR* is directly repressed by ER and its upregulation promotes ligand-independent ER activities and contributes to tamoxifen resistance.

Oncogene advance online publication, 14 September 2015; doi:10.1038/onc.2015.340

INTRODUCTION

Long non-coding RNAs (lncRNAs) are a major class of newly identified non-coding transcripts that are usually composed of more than 200 nucleotides. Accumulating evidence suggests that lncRNAs play critical roles in regulating a wide range of cellular processes by affecting various aspects of protein, DNA, and RNA expression and interactions.[1–5] Large-scale RNA sequencing (RNA-seq) studies have revealed that lncRNAs are abundantly transcribed from the genome; a recent study comprehensively examined over 7000 RNA-seq libraries and uncovered nearly 60,000 lncRNAs from the human transcriptome.[6] Out of numerous cancer-associated lncRNAs, *HOTAIR* (HOX antisense intergenic RNA) was among the most upregulated in breast cancer. Localized in chromosome 12, *HOTAIR* is 2.2 kb in length and transcribed from the antisense strand of the *HOXC* locus. It has been shown to interact with polycomb repressive complex 2 to reprogram the chromatin state and induce cancer metastasis.[7,8] *In vivo* experiments showed that *HOTAIR* is sufficient and required to promote invasion of breast carcinoma cells.[7] Concordantly, *HOTAIR* and EZH2 expression levels were highly correlated in breast cancer tissues and high *HOTAIR* level is associated with worse prognosis.[9,10] In addition, these studies reported that strong HOTAIR expression correlated with estrogen receptor (ER) and PR positivity, and that *HOTAIR* expression is a strong predictor of poor clinical outcome especially in ER-positive breast cancer.[9,10]

These results provide first lines of evidence that the lncRNA *HOTAIR* may play important roles in regulating breast cancer progression. Tamoxifen, an antagonist of the ER, is the most commonly used treatment for ER-positive breast cancer. Despite great success in improving the overall survival of breast cancer patients, development of tamoxifen resistance (TamR) is persistently seen in the clinic and is a major cause of breast cancer recurrence and mortality.[11] Understanding the biological mechanisms underlying this acquired resistance to tamoxifen is thus of substantial clinical significance.[12]

ER is a hormonal transcription factor that is liganded and activated by estrogen. ER regulates target genes that control endocrine response and cell cycle progression.[5,13,14] Tamoxifen competes with estrogen for binding to the ER protein, thereby inhibiting the conventional ER transcriptional program.[5,14,15] Using chromatin immunoprecipitation sequencing (ChIP-seq), a recent study has mapped genome-wide ER binding profiles in primary breast cancers and found that ER is still recruited to the chromatin in TamR breast cancer, but to new regulatory regions associated with poor clinical outcome.[16] This aberrant ER transcriptional activity is proposed to be regulated by various oncogenic mechanisms and have critical functions in mediating tamoxifen resistance and tumor progression. Here we report that *HOTAIR* is overexpressed in TamR breast cancer. It directly interacts with the ER protein to enhance ER transcriptional activity and thus ligand-independent breast cancer growth. Our study will not only inform about the mechanistic underpinnings of breast cancer progression but also provide evidence supporting therapeutic potentials of lncRNA targeting in breast cancer treatment.

RESULTS

HOTAIR is upregulated in tamoxifen-resistant, ER-positive breast cancer

To determine lncRNAs that may contribute to breast cancer tamoxifen resistance, we re-analyzed publicly available data set

[1]Division of Thoracic Surgery, Cancer Center of Guangzhou Medical University, Guangzhou, Guangdong, China; [2]Division of Hematology/Oncology, Department of Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA and [3]Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. Correspondence: Dr J Yu, Division of Hematology/Oncology, Department of Medicine, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, 303 E. Superior St. Lurie 5-117, Chicago, IL 60611, USA.
E-mail: jindan-yu@northwestern.edu
[4]These authors contributed equally to this work.
Received 8 May 2015; revised 23 July 2015; accepted 28 July 2015

profiling gene expression in wild-type MCF7 cells as well as its TamR derivatives treated with ethanol or 17β-estradiol (E2) for 4 h (GSE5840).[17] Our analysis revealed 37 lncRNA genes that were repressed by estrogen and became upregulated in TamR cells (Figure 1a). Among the top deregulated lncRNAs are *HOTAIR* and TP53TG1. Although *HOTAIR* has been shown to be upregulated in metastatic breast cancer,[7,10] its role in TamR has not been investigated. To examine this, we performed *in situ* hybridization to probe the abundance of *HOTAIR* lncRNA in breast cancer tissues, comparing between matched primary and TamR breast carcinoma samples. Our results showed that *HOTAIR* localized primarily in the nuclei but was also present in the cytoplasm (Figure 1b). Most primary breast cancer tissues had weak *HOTAIR* staining, whereas TamR breast cancer generally exhibited moderate to strong *HOTAIR* staining. Overall, *HOTAIR* expression level was significantly higher in TamR breast cancer than primary, hormone-naive tumors (Figure 1c). Being consistent with this, quantitative reverse transcriptase – PCR analysis showed that tamoxifen treatment for 7 days significantly increased *HOTAIR* lncRNA levels in both MCF7 and T47D cells, while dramatically decreasing the expression of *GREB1*, a known ER-induced gene (Figures 1d and e). As tamoxifen is known to compete with estrogen to inhibit estrogen-induced ER activities, next we examined whether *HOTAIR* is a target of ER-mediated transcriptional regulation.

The lncRNA *HOTAIR* is directly repressed by estrogen receptor

To examine whether estrogen regulates *HOTAIR* expression, we carried out quantitative reverse transcriptase – PCR analysis of MCF7 cells treated with increasing doses of E2. *HOTAIR* expression was greatly inhibited for up to sevenfold, while *GREB1* was increased as expected (Figure 2a and Supplementary Figure S1A). Estrogen inhibited *HOTAIR* expression in a dose- and time-dependent manner (Figures 2a and b). *HOTAIR* level was decreased about twofold after 4 h of E2 treatment and nearly 10-fold after 24 h of E2 treatment, whereas *GREB1* was gradually induced by around 20-fold at 4 h and reached a plateau of >30-fold after 8 h (Supplementary Figure S1B). A similar trend of inhibition of *HOTAIR* expression by estrogen was observed in a different ER+ breast cancer cell line T47D, despite T47D being much less responsive to estrogen as indicated by much less *GREB1* induction (Figures 2c and d). Furthermore, *HOTAIR* level is considerably restored in breast cancer cells following hormone deprivation, wherein *GREB1* expression was lost (Figure 2e). Next, to determine whether estrogen inhibits *HOTAIR* expression through direct ER binding to *HOTAIR* regulatory elements, we re-analyzed a previously published study involving an ER ChIP-seq data set that was performed in MCF7 cells (GSE23893).[18] We observed a very strong ER binding site at a genomic region about 14.5 kb upstream to the transcription start site of the *HOTAIR* gene (Figure 2f). In addition, this region is strongly occupied by H3K4me1 and H3K27ac (GSE40129), supporting its being an active enhancer (Supplementary Figure S2A). ER ChIP followed by quantitative reverse transcriptase – PCR analysis confirmed that estrogen stimulation significantly increased ER binding to this region as well as to positive control gene *GREB1*, but not to the negative control gene *KIAA0066* (Figure 2g). Further, chromosome conformation capture (3C) experiment demonstrated estrogen-induced DNA looping between the transcription start site of the *HOTAIR* gene (anchor primer) and

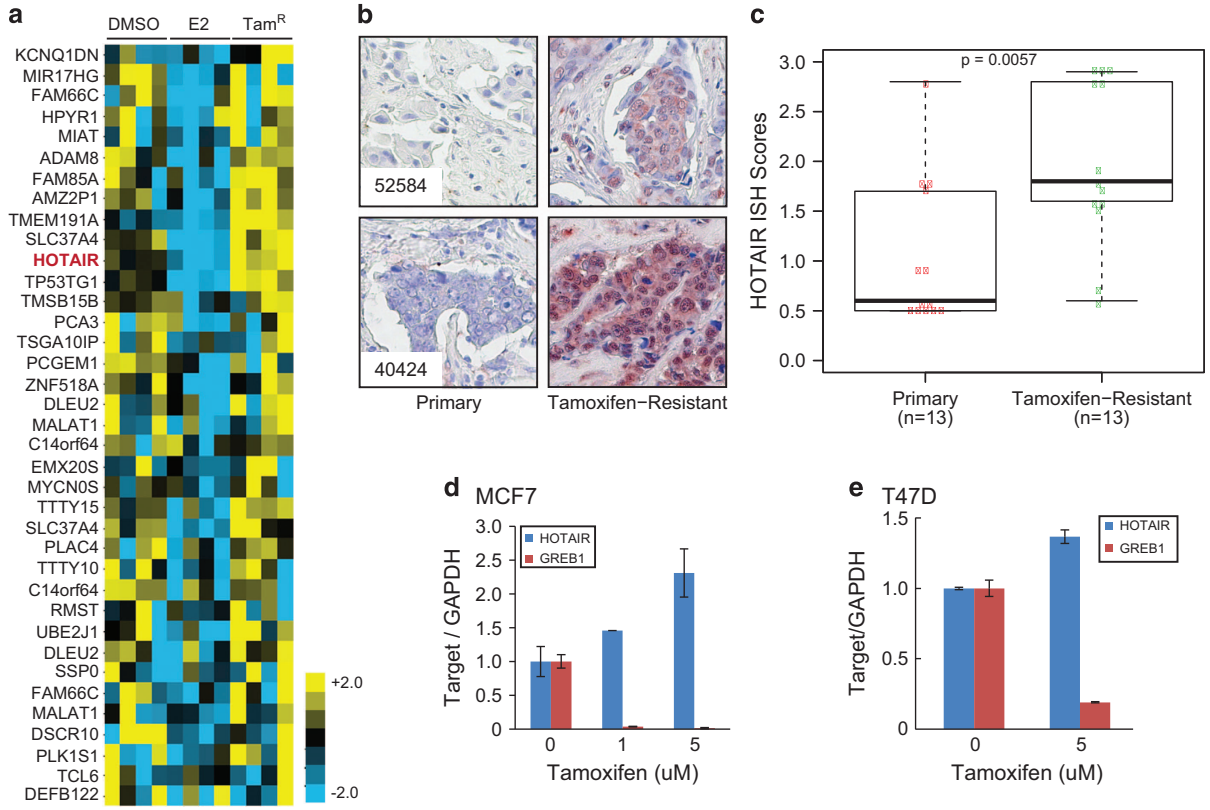


Figure 1. *HOTAIR* is upregulated in tamoxifen-resistant (TamR) breast cancer. (**a**) Heat map showing lncRNAs that are repressed by estradiol but upregulated in TamR MCF7 cells. Microarray data were downloaded from GEO with GSE5840 and re-analyzed for lncRNA expression. *HOTAIR* is shown in red. (**b**) *HOTAIR in situ* hybridization (ISH) staining in two representative pairs of primary and TamR breast cancers. (**c**) Boxplot showing *HOTAIR* ISH staining intensity in a set ($n = 13$) of matched primary and TamR breast cancers. (**d, e**) *HOTAIR* expression is increased by tamoxifen treatment. Quantitative reverse transcriptase – PCR analysis of *HOTAIR* and *GREB1* was done in MCF7 (**d**) and T47D (**e**) cells treated with increasing doses of tamoxifen for 7 days. Gene expression was normalized to GAPDH. Data shown are mean ± s.e.m. and are representative of at least two independent experiments.

© 2015 Macmillan Publishers Limited

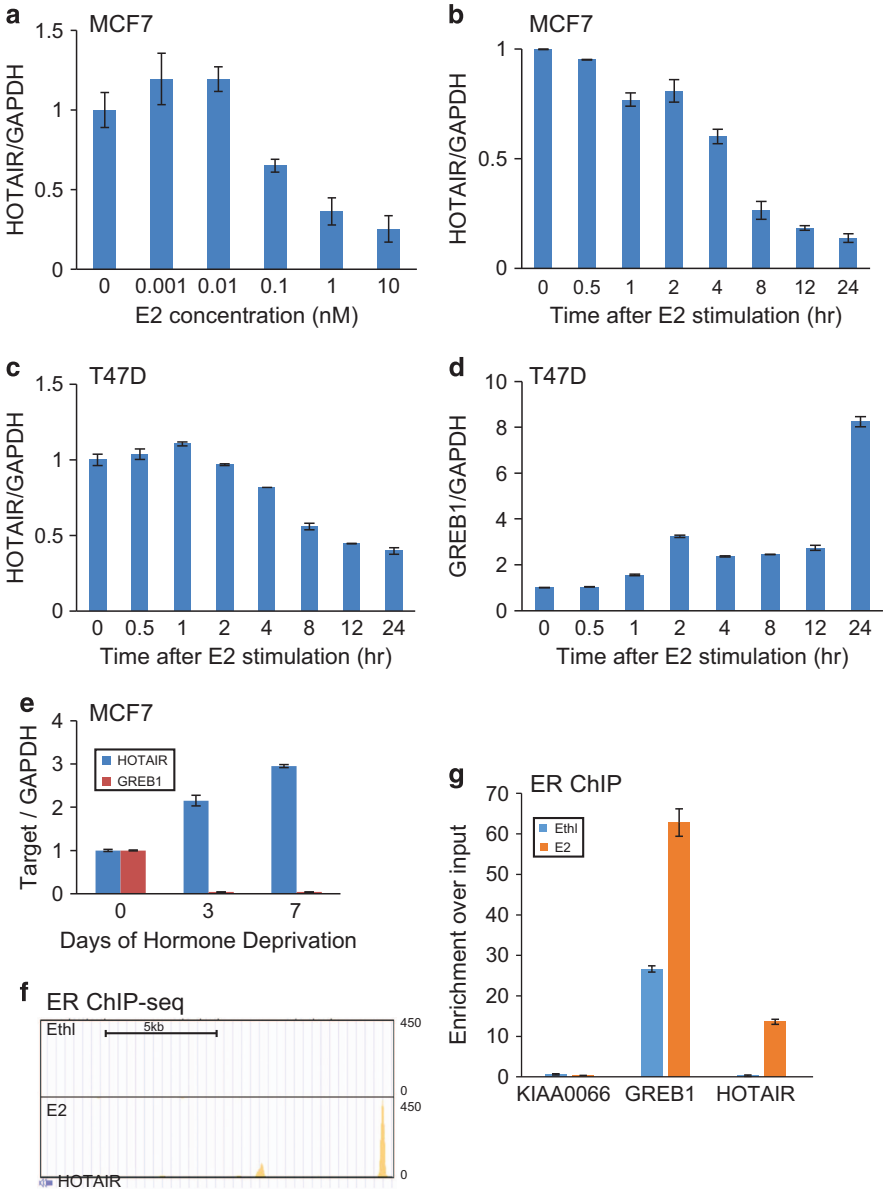


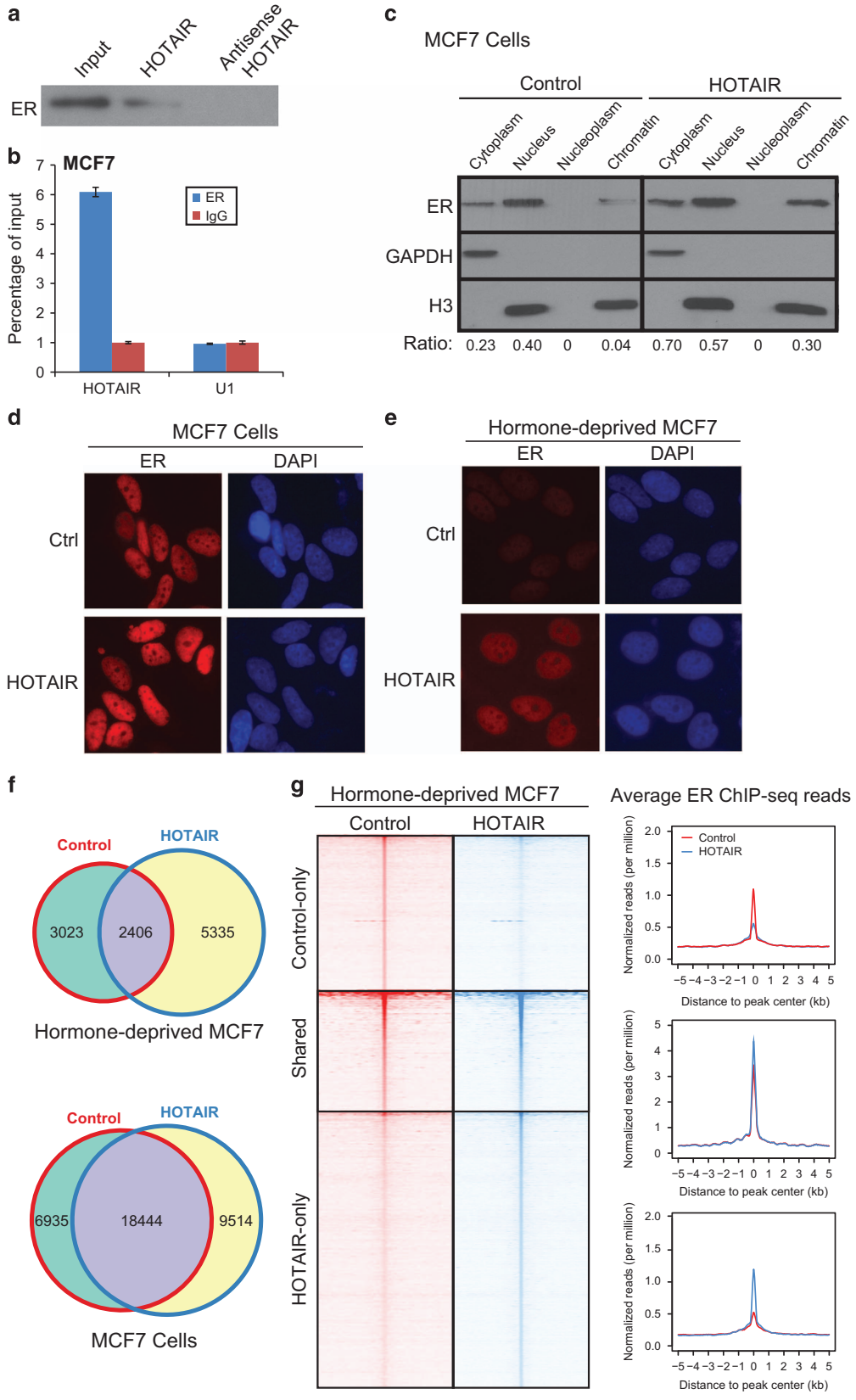
Figure 2. The lncRNA *HOTAIR* is directly repressed by estrogen through the estrogen receptor (ER). (**a**) Estrogen inhibits *HOTAIR* expression in a dose-dependent manner. MCF7 cells were hormone starved for 3 days and treated with increasing amounts of estradiol (E2) for 6 h. RNAs were then collected and subjected to quantitative reverse transcriptase – PCR (qRT – PCR) analysis of gene expression. (**b**) Estrogen inhibits *HOTAIR* expression in a time-dependent manner. MCF7 cells were treated with 1 nM E2 and collected at different time points for gene expression analysis by qRT – PCR. (**c, d**) Estrogen inhibits *HOTAIR* expression in T47D cells. T47D cells were hormone deprived for 3 days followed by E2 stimulation for up to 24 h. Cells were then collected at different time-points for qRT – PCR analysis of *HOTAIR* and *GREB1* expression and normalized to GAPDH. (**e**) Estrogen depletion restores *HOTAIR* level. qRT – PCR analysis of *HOTAIR* and *GREB1* in MCF7 cells subjected for hormone deprivation for up to 7 days. (**f**) Genome Browser view of ER binding events at the enhancer of the *HOTAIR* gene. ER chromatin immunoprecipitation sequencing (ChIP-seq) was performed in MCF7 cells stimulated with ethanol (Ethl) or E2[18] and re-analyzed using HOMER (GSE23893). (**g**) ChIP – qPCR showing ER binding to the *HOTAIR* distal enhancer. ER ChIP was performed in hormone-starved MCF7 cells stimulated with ethanol (vehicle) or E2 for 45 min. Enrichment of ER at specific genomic regions including *GREB1* and *HOTAIR* enhancers was evaluated by qPCR. The *KIAA0066* gene was utilized as a negative control as previously described.[34] qPCR, quantitative PCR.

the ER-bound enhancer (P4; Supplementary Figure S2B). Taken together, our data showed that *HOTAIR* is directly repressed by estrogen and is therefore upregulated following hormone deprivation and in TamR breast cancer.

HOTAIR directly interacts with ER and enhances ER transcriptional activities

Next, we asked what is the role of elevated HOTAIR in breast cancer ER signaling and tamoxifen resistance. *HOTAIR* has been previously shown to directly interact with chromatin-modifying proteins such as EZH2 and LSD.[15,19,20] As *HOTAIR* is upregulated in TamR breast cancer cells which often have altered ER program,[16] we asked whether *HOTAIR* might regulate ER function. This may shed light on the mechanisms underlying recently reported correlation between *HOTAIR* expression and ER positivity in primary specimens.[9,10] To test this, we first examined whether *HOTAIR* lncRNA could physically interact with the ER protein using RNA pull-down assay. Briefly, we carried out *in vitro* transcription to synthesize biotinylated RNA probes from sense and antisense *HOTAIR* DNA templates, which were then incubated with MCF7

© 2015 Macmillan Publishers Limited



nuclear extracts to allow protein – RNA interactions and precipitated, along with its interacting proteins, with streptavidin beads. Western blot analysis demonstrated that the sense HOTAIR RNA probe, but not the antisense transcript, pulled down the ER protein (Figure 3a). On the other hand, we carried out RNA immunoprecipitation assay and found that the ER antibody

© 2015 Macmillan Publishers Limited

significantly enriched for *HOTAIR*, as opposed to IgG control, whereas the negative control RNA U1 did not exhibit differential enrichment (Figure 3b).

Subsequently, we inquired into the consequences of *HOTAIR*–ER interaction, in order to speculate how ER activities may be affected as a result. By separating MCF7 cell lysates into cytoplasmic, nuclear, nucleoplasm and chromatin-bound fractions, we observed that ER, as expected, localized primarily within the nucleus as opposed to cytoplasm. HOTAIR overexpression substantially increased ER protein levels, suggesting potential roles of *HOTAIR* in enhancing ER transcriptional functions (Figure 3c and Supplementary Figure S3). Moreover, immunofluorescent staining confirmed noticeable increase of nuclear ER following *HOTAIR* overexpression (Figure 3d). Interestingly, this *HOTAIR*-mediated increase in nuclear ER level was also true under hormone-starved condition, in which there is only minimal estrogen present to activate ER translocation into the nucleus, suggesting the roles of *HOTAIR* in enhancing ligand-independent ER function (Figure 3e). To confirm the notion that *HOTAIR* may augment ER genomic targeting, we conducted ER ChIP-seq in MCF7 cells grown in the presence and absence of estrogen. As expected, the total number of ER binding sites was 4.6-fold higher in estrogen-stimulated vs hormone-deprived cells (Figure 3f). Importantly, upon overexpression of *HOTAIR*, global ER binding events were greatly increased under both conditions. Heat map and average intensity analysis of the various groups of ER peaks demonstrated a clear increase in ER ChIP-seq read intensity in both shared and *HOTAIR*-only groups, representing a majority of the ER binding events (Figure 3g and Supplementary Figure S4A). This *HOTAIR*-mediated increase of ER binding events was more prominent in the absence of estrogen, suggesting important functions of *HOTAIR* in regulating ligand-independent ER activities. Concordantly, quantitative PCR analysis of several previously reported ER target genes, such as *GREB1, TFF1, PR*, and *CTSD*, demonstrated that *HOTAIR* overexpression significantly increased ER occupancy at most of these genes (Supplementary Figure S4B). Similarly, the increase in ER binding at target genes was more prominent in hormone-deprived MCF7 cells (Supplementary Figure S4C). Next, we proceeded to investigate to what extent *HOTAIR* impacts ER-mediated transcriptional activities particularly in a hormone-deprived environment.

HOTAIR drives estrogen-independent ER transcriptional program

To identify *HOTAIR*- and estrogen-regulated genes, we conducted microarray profiling of hormone-deprived and estrogen-stimulated MCF7 cells with control or *HOTAIR* overexpression. Data analysis identified 132 and 112 genes that were induced and repressed by *HOTAIR*, respectively. Importantly, hierarchical clustering followed by heat map view revealed that a majority of *HOTAIR*-induced genes are also induced by E2 stimulation, whereas *HOTAIR*-repressed genes tend to become downregulated by estrogen (Figure 4a). Concordantly, gene set enrichment analysis demonstrated that E2-induced genes were significantly upregulated following *HOTAIR* overexpression, even in the absence of estrogen, whereas E2-repressed genes were strongly downregulated by HOTAIR (Figures 4b and c). Gene ontology analysis showed that *HOTAIR*-induced genes were significantly enriched for response to protein stimulus and regulation of cell death and apoptosis, being consistent with the functions of estrogen-mediated ER signaling (Figure 4d; Supplementary Tables S1 and S2). To confirm *HOTAIR* regulation of ER-mediated transcriptional program, we performed quantitative reverse transcriptase – PCR analysis of several known ER-target genes such as *GREB1, TFF1*, and *c-MYC*. Indeed, our data showed that *HOTAIR* overexpression induced ER-target gene expression in the absence of estrogen and further potentiated the effects of E2 (Figures 4e and g). Taken together, we provide evidence for a model by which the lncRNA *HOTAIR* increases ER protein level and enhances its chromatin binding and thus the ER transcriptional program, even in an estrogen-depleted environment. As *HOTAIR* is upregulated in TamR breast cancer, we next asked whether *HOTAIR* contributes to the development of tamoxifen resistance in breast cancer, wherein tamoxifen abolishes estrogen-mediated activation of ER, similar to hormone starvation.

LncRNA *HOTAIR* promotes tamoxifen-resistant breast cancer progression

To determine the role of *HOTAIR* in breast cancer, we first overexpressed *HOTAIR* in MCF7 cells (Figure 5a). Cell proliferation assay showed that *HOTAIR* overexpression increased MCF7 cell growth (Figure 5b). On the other hand, *HOTAIR* knockdown in T47D cells markedly reduced cell proliferation (Figures 5c and d). To provide direct evidence linking *HOTAIR* to tamoxifen resistance, we generated a TamR MCF7 cell line by continuously culturing the cells in the presence of 5 μM tamoxifen for several months. Consistent with previous *HOTAIR* staining results in TamR breast tumors, *HOTAIR* level showed a remarkable fourfold increase following long-term treatment of tamoxifen (Figure 5e). To determine whether this upregulated *HOTAIR* is critical for the TamR MCF7 cell growth, we performed *HOTAIR* knockdown using two independent short hairpin RNA constructs (Figure 5f). Subsequently, we performed cell proliferation assay to investigate to what extent *HOTAIR* contributes to tamoxifen resistance. As demonstrated in Figure 5g, knockdown of *HOTAIR* significantly decreased TamR MCF7 cell growth, suggesting that tamoxifen resistance may be reverted by targeting or depleting *HOTAIR*. Consistently, clonogenic assays showed that *HOTAIR* knockdown greatly inhibited the colony-formation abilities of the TamR cells, further supporting the role of *HOTAIR* in mediating TamR cell growth (Figure 5h).

DISCUSSION

With the emergence of studies focusing on the functional attributes of nonprotein-coding RNA transcripts, such as lncRNAs, it has been revealed that these lncRNAs may contribute

Figure 3. *HOTAIR* interacts with the estrogen receptor (ER) protein and enhances ER genomic action. (**a**) *HOTAIR* lncRNA interacts with the ER protein. RNA pull-down assay was performed in MCF7 cells using biotin-labeled *HOTAIR* RNA probe transcribed *in vitro*. The antisense *HOTAIR* probe was used as negative control. (**b**) ER protein binds to *HOTAIR* lncRNA. MCF7 cells were subjected to RNA immunoprecipitation assay using an anti-ER antibody or IgG control. Immunoprecipitation-enriched RNA was then analyzed by quantitative reverse transcriptase – PCR. U1 RNA was utilized as a negative control. (**c**) *HOTAIR* overexpression increases ER protein level. MCF7 cell lysates were separated into the cytoplasm, nuclear, nucleoplasm, and chromatin-bound fractions and were detected by western blot analysis. GAPDH and H3 were utilized as loading controls for cytoplasmic and nuclear/chromatin fractions, respectively. Quantification was done by measuring band intensity with ImageJ (NIH, Bethesda, MD, USA) and normalizing to loading control. (**d, e**) Ectopic overexpression of *HOTAIR* increases nuclear ER level. ER immunostaining was performed in control and *HOTAIR*-overexpressing MCF7 cells grown in the presence (**d**) and absence (**e**) of estrogen. (**f**) Overlap of ER-binding sites detected by chromatin immunoprecipitation sequencing (ChIP-seq) in MCF7 cells with control or HOTAIR overexpression in the absence and presence of estrogen. (**g**) Heat map depicting ER ChIP-seq read intensity around (±5 kb) peak centers detected in control or HOTAIR-overexpressing MCF7 cells under hormone-starved condition. Average ER ChIP-seq read intensity around ER binding sites (±5 kb) is shown on the right.

© 2015 Macmillan Publishers Limited

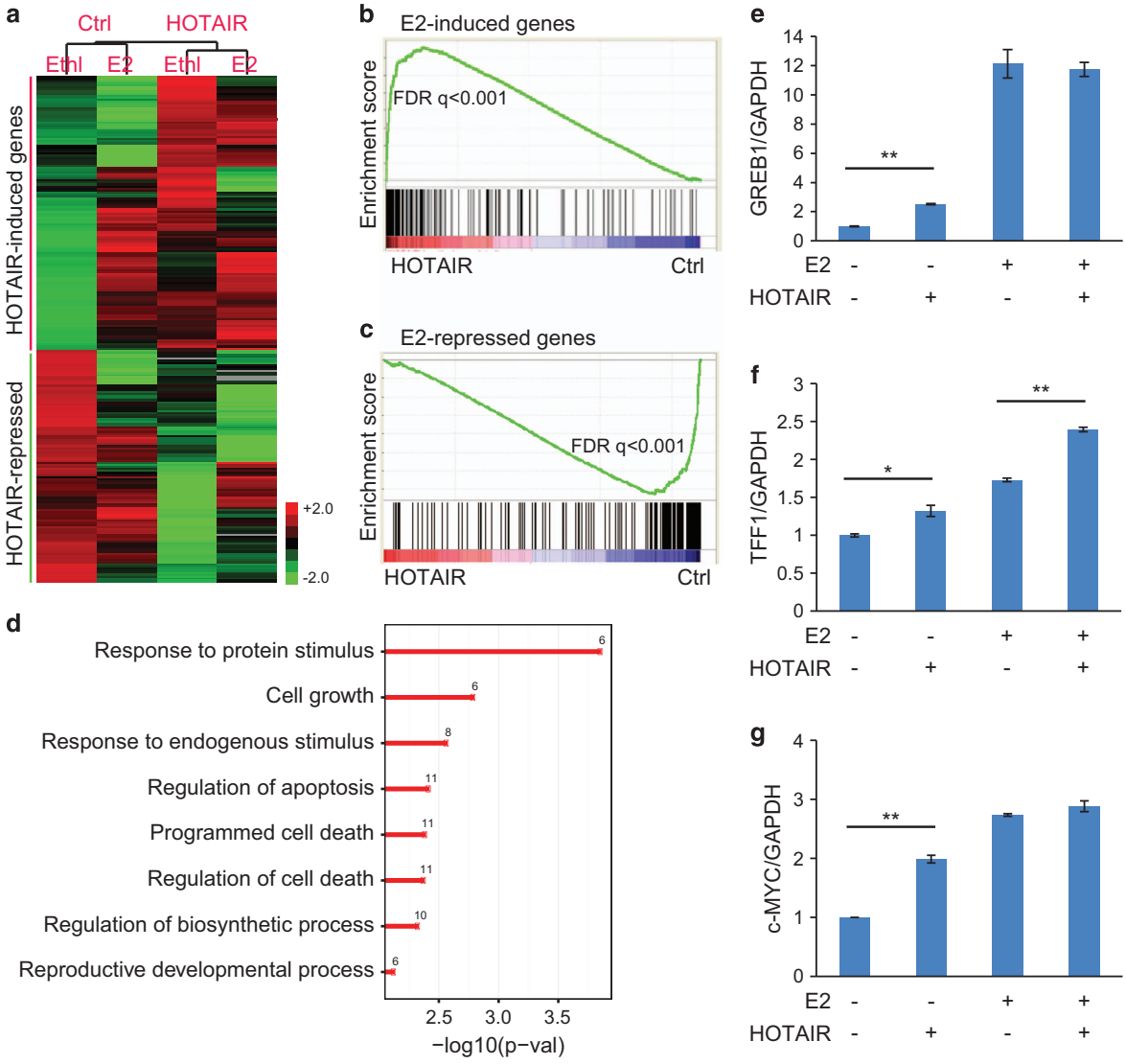


Figure 4. *HOTAIR* overexpression enhances estrogen receptor (ER) transcriptional program. (**a**) *HOTAIR*-induced and repressed genes are, respectively, increased and decreased by estrogen. Expression microarray was utilized to profile gene expression in control and *HOTAIR*-overexpressing MCF7 cells that were hormone-deprived for 3 days followed by either ethanol or 1nM estradiol (E2) treatment for 6 h. Expression of genes induced or repressed by *HOTAIR* for at least twofold were clustered and visualized using heat map. (**b, c**) Estrogen-induced genes are significantly enriched for upregulation by HOTAIR (**b**), whereas estrogen-repressed genes are downregulated by *HOTAIR* (**c**). GSEA was carried out to determine the enrichment of E2-induced and -repressed gene sets in the expression data set comparing control and *HOTAIR*-overexpressing MCF7 cells. (**d**) *HOTAIR*-induced genes are enriched for cell growth and response to protein stimulus. Gene ontology (GO) analysis was performed using 132 genes that are increased by *HOTAIR* by at least twofold. Shown here are representative GO terms that are significantly enriched ($P < 0.05$). (**e** – **g**) *HOTAIR* enhances ER-target gene expression. Quantitative reverse transcriptase – PCR analysis of representative ER-induced genes was performed in control and *HOTAIR*-overexpressing MCF7 cells stimulated with ethanol or estrogen. Data were normalized to GAPDH. Error bars: mean ± s.e.m. *$P < 0.05$ and **$P < 0.01$.

significantly to the biological processes involved in physiological as well as pathological conditions. Numerous lncRNAs have been identified as critical players during cancer development; some may be beneficial by acting as tumor or metastasis suppressors (for example, *GAS5*,[21] *MEG3*,[22] *LIFR*[23]), whereas others may be detrimental by promoting oncogenesis (for example, *PCA3* or, as previously named, *DD3*,[24] *PCAT-1*,[25] *SChLAP1*[26]). Previous studies have shown that lncRNAs exhibit great diversity in their functions and mechanisms of action, which include but are not limited to epigenetic transcriptional regulation, association with enhancer and chromosomal looping, and mRNA processing and translation.[27] Several unique properties of lncRNAs make them highly useful in the clinic, with potential utilities including their use as diagnostic biomarkers due to their tissue specificity,[25,28] as well as in lncRNA-based therapies by means of RNA interference.[29] Yet, lncRNAs have just begun to be identified and cataloged; a majority of them remain to be characterized.

Gupta *et al.*[7] reported in 2010 that the lncRNA *HOTAIR* is notably increased in primary breast tumors as well as during metastases. Specifically, by interacting with EZH2 of the polycomb repressive complex 2 complex, which catalyzes trimethylation at histone H3 lysine 27 (H3K27me3) and is upregulated in a variety of aggressive cancers, *HOTAIR* was demonstrated to alter chromatin structure and regulate gene expression, thereby giving rise to an invasive cancer phenotype. In this study, we provide experimental evidence that *HOTAIR* is also critically involved in conferring tamoxifen resistance to MCF7 cells, which represents a major challenge in the clinic

© 2015 Macmillan Publishers Limited

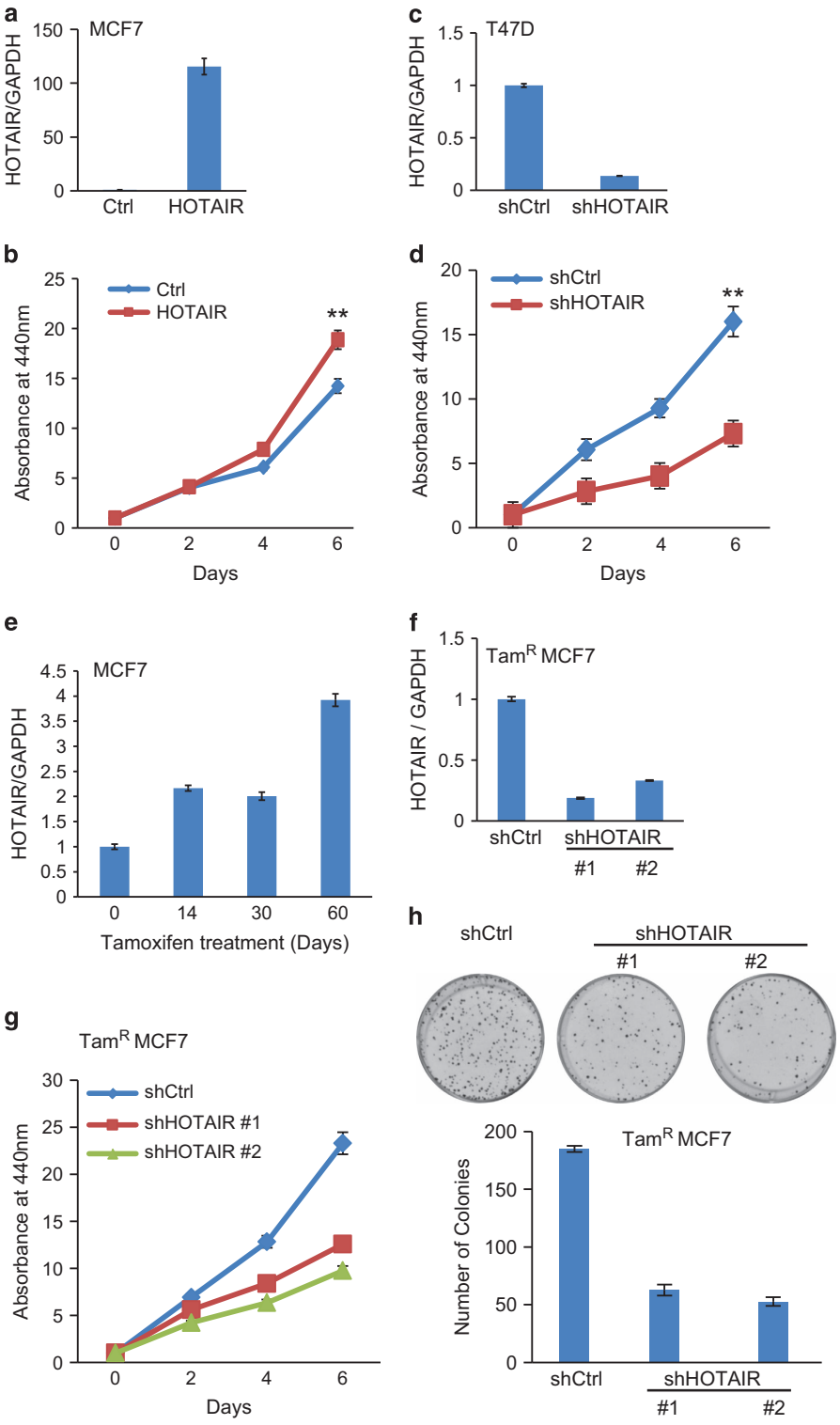


Figure 5. *HOTAIR* promotes breast cancer cell growth and tamoxifen resistance. (**a, b**) *HOTAIR* overexpression increases breast cancer cell growth. *HOTAIR* was overexpressed in MCF7 cells through lentiviral transduction, with overexpression confirmed by quantitative reverse transcriptase – PCR (qRT – PCR) (**a**). Cell proliferation was determined by WST-1 cell growth assay (**b**). Data shown are mean ± s.e.m. and are representative of at least two independent experiments, $^{**}P < 0.01$. (**c, d**) HOTAIR knockdown decreases breast cancer cell growth. HOTAIR was depleted in T47D cells through short hairpin RNA (shRNA) lentiviral transduction (**c**) and cell growth was evaluated by WST-1 assay (**d**). Data shown are mean ± s.e.m. and are representative of at least two independent experiments, $^{**}P < 0.01$. (**e**) *HOTAIR* is upregulated in tamoxifen-resistant (TamR) breast cancer cells. MCF7 cells were continuously grown in medium containing 5 μM tamoxifen and periodically harvested for RNA isolation and qRT – PCR analysis. (**f, g**) *HOTAIR* knockdown decreases TamR breast cancer cell growth. *HOTAIR* was depleted in TamR MCF7 cells through lentiviral transduction of two shRNA constructs. Gene expression was determined by qRT – PCR (**f**) and cell growth by WST-1 assay (**g**). (**h**) *HOTAIR* knockdown inhibits colony formation abilities of breast cancer cells. Colony formation assays and quantifications were performed in TamR MCF7 cells stably expressing control or *HOTAIR*-targeting shRNAs.

© 2015 Macmillan Publishers Limited

today. Tamoxifen, belonging to the class of selective ER modulators, is a competitive antagonist of ER that was developed in the 1970s and has been the mainstay treatment for ER-positive breast cancer, which accounts for at least 70% of all breast cancers.[12] Despite its initial success in reducing disease mortality and improving survival, tamoxifen therapy frequently led to the onset of resistance, and recurrence was reported to occur within 15 years in one-third of patients treated with tamoxifen.[12,30] Thus, it has become imperative to understand the mechanisms for acquisition of tamoxifen resistance and to develop targeted therapies to improve treatment for breast cancer.

Our results showed that *HOTAIR* is highly upregulated in the tumors of TamR breast cancer patients compared to their primary tumors before treatment. Moreover, physical interaction between *HOTAIR* and the nuclear hormone receptor ER was detected, which in turn resulted in significant amount of nuclear ER even under estrogen-depleted conditions, thus allowing ER genomic targeting and consequently inducing the ER transcriptional program. Importantly, this phenomenon of *HOTAIR*-mediated activation of ER function in the absence of estrogen indicated a potential route to ligand independence that is manifested in TamR cells. Furthermore, by generating a TamR MCF7 cell model, we showed that *HOTAIR* was consistently upregulated over long periods of drug treatment. In addition, we demonstrated that *HOTAIR* significantly contributes to the growth of these TamR cells. Therefore, in our present study we provide evidence for a novel mechanism that is employed by the lncRNA *HOTAIR* to promote ER activation in the absence of estrogen and drive tamoxifen resistance. Because of this crucial role *HOTAIR* plays in the progression of breast cancer and development of drug resistance, it holds great promise as a useful biomarker and potential therapeutic target.

MATERIALS AND METHODS

Patient specimens and cell lines

All breast cancer tissue specimens ($n = 13$) were collected via surgical resection or biopsy from patients diagnosed between January 2006 and February 2014 at the Cancer Center of Guangzhou Medical University. The study protocol was approved by the Ethics Committee of Cancer Center of Guangzhou Medical University. In general, with $n = 10$, for a continuous outcome, there will be $>89\%$ power to reject the null hypothesis of no difference when the difference is 1.5 s.d. or more, using two-sided *t*-test and a type 1 error of 0.05. MCF7 and T47D cell lines were ordered from ATCC.

Plasmids, reagents, quantitative PCR and western blotting

HOTAIR sequence was amplified by PCR and subsequently cloned into the expression vector pCDH-MSCV-mcs-EF1-GFP-T2A-Pu (SBI) at *Eco*R1 and *Not*1 sites using Cold Fusion kit (SBI). The shHOTAIR was cloned into the pLKO lentivirus system. All PCR primers for cloning are listed in Supplementary Table S3 and high-fidelity enzyme Phusion was used for PCR amplification. All PCR products were verified by DNA sequencing. Specific antibodies used in this work include rabbit ER (06-935, Millipore, Billerica, MA, USA), mouse ER (sc-8002, Santa Cruz, Dallas, TX, USA), mouse GAPDH (ab9484, Abcam, Cambridge, MA, USA) and rabbit H3 (ab1791, Abcam). Other reagents include beta-estradiol (E8875, Sigma, St Louis, MO, USA) and 4-hydroxytamoxifen (H6278, Sigma-Aldrich, St Louis, MO, USA). All primers were designed using primer 3 and synthesized by Integrated DNA Technologies (Coralville, IA, USA; Supplementary Table 3). Quantitative PCR was performed using SYBR Green by StepOne Plus in three technical replicates and significance was determined by two-sided *t*-tests. Each experiment was repeated independently at least two times. Western blotting was carried out using standard protocol and repeated at least two times. Band intensity on western blot was quantified with ImageJ and normalized to each respective control to obtain the ratio of ER protein level.

3C assay

The digestion map of commonly used restriction enzymes around the enhancer/promoter region of HOTAIR locus (from −103 to +83 kb) and *Bgl*II was selected for digestion, as *Bgl*II sites show a distribution that will enable appropriate primers to be designed to generate 200 – 350 bp PCR products on re-ligation. All primers are designed based on the forward strand immediately upstream of a *Bgl*II restriction site (Supplementary Figure S2B and Supplementary Table S3). 3C experiments were conducted according to the standard 3C protocol as previously described.[31] Briefly, fixed chromatin of hormone-starved or E2-treated MCF7 cells (1×10^7) was digested with *Bgl*II overnight and incubated with 50 units of T4 DNA ligase (10799009001, Sigma-Aldrich) overnight in a volume of 7 ml to keep the DNA concentration at 2 – 3 ng/ml to favor intramolecular ligation.

RNA pull-down assay

RNA pull-down was performed as previously described.[20] Briefly, biotin-labeled RNAs were transcribed from DNA templates with biotin-UTP, NTP mix, and T7 RNA polymerase (Promega, Madison, WI, USA), treated with RNase-free DNase I (Promega) and purified with RNeasy Mini kit (QIAGEN). Nuclei were extracted from MCF7 cells and resuspended in 1 ml RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM dithiothreitol, 0.5% NP40, 1 mM phenylmethylsulfonyl fluoride, and protease inhibitor (Roche Complete Protease Inhibitor Cocktail Tablets)), and subsequently subjected to mechanical shearing using a dounce homogenizer. For precipitation assays, fragmented nuclear extract and the RNA probe were incubated at room temperature (RT) for 60 min, and 60 µl of Streptavidin agarose beads (Invitrogen, Grand Island, NY, USA) were added to each binding reaction and further incubated at RT for 1 h. After five times of washing with PBS, samples were boiled in SDS buffer and subjected to western blot analysis.

RNA immunoprecipitation

RIP protocol was derived from published reports.[20] Briefly, cells were treated with 0.3% formaldehyde for 10 min at 37 °C, then added with glycine to a final concentration of 0.125 M and then incubated at RT for 5 min. Cells were then washed twice in cold phosphate buffered saline and pellet was resuspended in 1 ml of RIPA buffer, which was incubated on ice with frequent vortexing for 30 min. Finally, the nuclear lysate was obtained by centrifugation at 13,000 r.p.m. for 10 min. To obtain bead and antibody complex, 20 µl protein beads were mixed with 1 µg antibody and rotated for 4 h at 4 °C. The complex was added to nuclear lysates and incubated overnight at 4 °C and then incubated with RNase-free DNase I (Promega) at 37 °C for 15 min and proteinase K at 45 °C for 45 min. Lastly, RNA was extracted with 1 ml TRIzol (Invitrogen) and analyzed by quantitative PCR.

Immunofluorescent staining

Cells were fixed with 4% formaldehyde for 15 min at RT and then permeabilized in 0.1% Triton X-100 for 15 min at RT. Cells were then washed by PBS for three times, followed by incubation with 5% normal goat serum for 30 min at RT. Subsequently, cells were incubated with primary antibody, the anti-mouse ER antibody (Santa Cruz), for 1 h at RT. After washing three times with PBS, cells were incubated with secondary antibody, Alexa Fluor 594 goat anti-mouse IgG (Invitrogen), for 1 h at RT. Finally, cells were washed three times with PBS and mounted using Prolong Gold Antifade Reagent (Invitrogen).

LncRNA *in situ* hybridization

Biotin-labeled antisense HOTAIR RNAprobe /5Biosg/G+C+C+TTGCTCCCTT +G+CCTGCATTTCT+C+T+G was synthesized by EXIQON (Woburn, MA, USA). For paraffin-embedded tissue, after deparaffinization and rehydration, the samples were treated with peroxidase-quenching solution; proteinase K was added to digest tissues before prehybridization and hybridization, which were carried out at 56 °C for 30 min and 4 h, respectively. Then streptavidin–horseradish peroxidase was used to react with the bound biotin-labeled probe. The signal was further amplified using TSA amplification kit (PerkinElmer, Waltham, MA, USA). Finally, the signal was revealed with Ultra Vision One polymer and aminoethyl carbazole chromogen (Thermo Fisher Scientific, Waltham, MA, USA). The stainings were then scored by eye by two pathologists, on a three-tiered scoring system, using the following criteria for the three tiered system:

© 2015 Macmillan Publishers Limited

0 = negative, 1 = equivocal/uninterpretable, 2 = weak positive and 3 = strong positive.

Gene expression microarray and data analysis

Total RNAs were isolated using TRIzol reagent (Invitrogen). The integrity of the RNA was verified using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Microarray profiling was performed using HumanHT-12 v 4.0 Expression BeadChip (Illumina, San Diego, CA, USA). Bead-level data were processed using GenomeStudio (Illumina), and the expression values were quantile-normalized using the limma package in Bioconductor.[32]

Genes having at least twofold changes in *HOTAIR*-overexpressing ells compared with the control cells in the absence of estrogen were defined as *HOTAIR*-regulated gene set. Genes with at least twofold change between ethanol and estrogen-stimulated MCF7 cells were defined as estrogen-regulated genes. Gene ontology terms enrichment was analyzed using DAVID 6.7.[33] Gene set enrichment analysis was performed as previously described.[34]

ChIP and ChIP-seq

ChIP experiments were carried out as previously described.[35] Antibodies used are Rabbit ER (06-935, Millipore) and Rabbit IgG (sc-2027, Santa Cruz). ChIP-quantitative PCR enrichment of target loci was normalized to input DNA and reported as % input ± s.e.m. ChIP DNA was prepared into libraries according to standard protocols using Bioo Scientific's DNA Sample Kit (cat. no. 514101, Austin, TX, USA). Libraries were sequenced using Illumina Hi-Seq platforms. Sequence reads were aligned to the Human Reference Genome (assembly hg19) using Burrows – Wheeler alignment tool (bwa) version 0.6.1.[36] Microarray and short-read sequencing data have been deposited in the GEO database with the accession number GSE61270.

Cell proliferation and clonogenic assay

Cell proliferation assay was carried out using the WST-1 kit according to the manufacturer's instruction (Clontech, Mountain View, CA, USA). Briefly, 5000 cells were seeded in a 24-well plate. After adding 50 µl WST-1 reagents per well, cultures were incubated for 2 h and the absorbance at a wavelength of 440 nm was determined using a microplate reader. For clonogenic assay, 500 cells were plated in each well of a six-well plate. When there was visible colony by naked eye, cells were fixed with 4% formaldehyde and were stained with crystal violet (0.25%). Colonies were then counted.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Bioinformatic analysis was supported by the computational resources and staff contributions provided for the Quest high-performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This work was supported by the the US Department of Defense W81XWH-13-1-0319 (to JY) and the Research Scholar Award RSG-12-085-01 (to JY) from the American Cancer Society. JK was supported in part by the National Institutes of Health Training Program in Oncogenesis and Developmental Biology (T32 CA080621), and YAY was supported in part by the National Institutes of Health/National Cancer Institute training grant T32 CA009560.

REFERENCES

1 Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009; **10**: 155–159.
2 Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009; **136**: 629–641.
3 Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011; **43**: 904–914.
4 Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009; **23**: 1494–1504.
5 Yoon JH, Abdelmohsen K, Kim J, Yang X, Martindale JL, Tominaga-Yamanaka K *et al.* Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination. *Nat Commun* 2013; **4**: 2939.
6 Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015; **47**: 199–208.
7 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; **464**: 1071–1076.
8 Gupta S, Iljin K, Sara H, Mpindi JP, Mirtti T, Vainio P *et al.* FZD4 as a mediator of ERG oncogene-induced WNT signaling and epithelial-to-mesenchymal transition in human prostate cancer cells. *Cancer Res* 2010; **70**: 6735–6745.
9 Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY, West RB. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PLoS One* 2012; **7**: e47998.
10 Sorensen KP, Thomassen M, Tan Q, Bak M, Cold S, Burton M *et al.* Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. *Breast Cancer Res Treat* 2013; **142**: 529–536.
11 Ring A, Dowsett M. Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* 2004; **11**: 643–658.
12 Musgrove EA, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* 2009; **9**: 631–643.
13 Doisneau-Sixou SF, Sergio CM, Carroll JS, Hui R, Musgrove EA, Sutherland RL. Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells. *Endocr Relat Cancer* 2003; **10**: 179–186.
14 Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 2000; **103**: 843–852.
15 Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 1998; **95**: 927–937.
16 Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012; **481**: 389–393.
17 Fan M, Yan PS, Hartman-Frey C, Chen L, Paik H, Oyer SL *et al.* Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. *Cancer Res* 2006; **66**: 11954–11966.
18 Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L *et al.* Integrative model of genomic factors for determining binding site selection by estrogen receptor--alpha. *Mol Syst Biol* 2010; **6**: 456.
19 Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA *et al.* Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev* 2010; **24**: 2615–2620.
20 Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F *et al.* Long non-coding RNA as modular scaffold of histone modification complexes. *Science* 2010; **329**: 689–693.
21 Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 2009; **28**: 195–208.
22 Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R *et al.* Activation of p53 by MEG3 non-coding RNA. *J Biol Chem* 2007; **282**: 24731–24742.
23 Chen D, Sun Y, Wei Y, Zhang P, Rezaeian AH, Teruya-Feldstein J *et al.* LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat Med* 2012; **18**: 1511–1517.
24 Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA *et al.* DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 1999; **59**: 5975–5979.
25 Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* 2011; **29**: 742–749.
26 Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* 2013; **45**: 1392–1398.
27 Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discovery* 2011; **1**: 391–407.
28 Lee GL, Dobi A, Srivastava S. Prostate cancer: diagnostic performance of the PCA3 urine test. *Nat Rev Urol* 2011; **8**: 123–124.
29 Davis ME, Zuckerman JE, Choi CH, Seligson D, Tolcher A, Alabi CA *et al.* Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* 2010; **464**: 1067–1070.
30 Clarke R, Leonessa F, Welch JN, Skaar TC. Cellular and molecular pharmacology of antiestrogen action and resistance. *Pharmacol Rev* 2001; **53**: 25–71.

© 2015 Macmillan Publishers Limited

31 Wu L, Runkle C, Jin HJ, Yu J, Li J, Yang X *et al*. CCN3/NOV gene expression in human prostate cancer is directly suppressed by the androgen receptor. *Oncogene* 2014; **33**: 504–513.
32 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi WX *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47.
33 Huang, da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
34 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
35 Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X *et al*. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 2010; **17**: 443–454.
36 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009; **25**: 1754–1760.

Supplementary Information accompanies this paper on the Oncogene website (http://www.nature.com/onc)

© 2015 Macmillan Publishers Limited

FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function

Yeqing Angela Yang[1,*], Jonathan C. Zhao[1,*], Ka-wing Fong[1], Jung Kim[1], Shangze Li[1], Chunxiao Song[2], Bing Song[1], Bin Zheng[1], Chuan He[2,3], Jindan Yu[1,4]

[1]Division of Hematology/Oncology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA;

[2]Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, the University of Chicago, Chicago, Illinois 60637, USA;

[3]Howard Hughes Medical Institute, the University of Chicago, Chicago, Illinois 60637, USA;

[4]Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Running Title: Feed-forward regulation between FOXA1 and TET1

Address correspondence and requests for reprints to:
Jindan Yu, M.D., Ph.D.
Division of Hematology/Oncology, Department of Medicine
Robert H. Lurie Comprehensive Cancer Center
Northwestern University, Feinberg School of Medicine
303 E. Superior St. Lurie 5-117
Chicago, IL 60611
Phone: 312-503-1761
Fax: 312-503-0189
E-mail: jindan-yu@northwestern.edu

SUMMARY

FOXA1 is a FKHD family protein that translates epigenetic signatures at target enhancers to lineage-specific transcription and differentiation. Through genome-wide location analyses, here we show that FOXA1 expression and occupancy are, in turn, required for the maintenance of these epigenetic signatures, namely DNA hypomethylation and histone 3 lysine 4 methylation. Mechanistically, this involves TET1, a 5-methylcytosine dioxygenase. We found that FOXA1 induces TET1 expression via direct binding to its cis-regulatory elements. Further, FOXA1 physically interacts with the TET1 protein through its CXXC domain. TET1 thus co-occupies FOXA1-dependent enhancers and mediates local DNA demethylation and concomitant histone 3 lysine 4 methylation, further potentiating FOXA1 recruitment. Consequently, FOXA1 binding events are markedly reduced following TET1 depletion. Together, our results support that FOXA1 is not only a reader, but also a writer, of the epigenetic signatures at lineage-specific enhancers and that TET1 and FOXA1 form a feed-forward regulatory loop for enhancer activation.

Highlights

- FOXA1 contributes to epigenetic activation of lineage-specific enhancers.
- *TET1* is a direct target of FOXA1-mediated transcriptional regulation.
- TET1 interacts with the FKHD domain of the FOXA1 protein through its CXXC domain.
- TET1 facilitates FOXA1 recruitment to target enhancers via DNA demethylation**.**

INTRODUCTION

Forkhead box A1 (FOXA1; also known as hepatocyte nuclear factor 3 alpha, or HNF3A) belongs to the forkhead family of transcription factors which have been well-known to play a pivotal role for the postnatal development of the mammary and prostate glands (Bernardo and Keri, 2012). FOXA1 is critical in directing hormone receptor-dependent transcriptional programs to regulate prostate- or breast-specific gene expression and cell differentiation (Bernardo et al., 2010; Gao et al., 2005). FOXA1 acts as a "pioneer transcription factor", whose function is to recognize lineage-specific enhancers and facilitate the recruitment of nuclear receptors to these sites (Jozwik and Carroll, 2012). Genome-wide location analyses have reported that FOXA1 preferentially recognizes and binds lineage-specific enhancers that are demarcated by active histone modifications including histone H3 lysine 4 mono- and di-methylation (H3K4me1, me2) (Lupien et al., 2008), histone 27 acetylation (H3K27ac) (Wang et al., 2012), as well as local DNA hypomethylation (Serandour et al., 2011). On the other hand, enforced expression of FOXA1 and its binding to DNA further induces H3K4 methylation and leads to DNA demethylation, forming a negative feedback loop (Serandour et al., 2011). However, the molecular mechanisms by which FOXA1 imposes this epigenetic switch have not been characterized.

TET (ten-eleven translocation) proteins are a family of DNA hydroxylases that oxidize the methyl group at the C5 position of methylated cytosine, enzymatically converting 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) in a sequential and iterative manner, ultimately leading to the removal of DNA methylation (Ito et al., 2011; Tahiliani et al., 2009). Through catalyzing DNA demethylation, TET proteins play important roles in embryonic stem cell maintenance and in

regulating appropriate lineage differentiation of these cells. These activities can be linked to the ability of DNA demethylation in modulating transcription factor occupancy and vice versa (Feldmann et al., 2013; Maurano et al., 2015). During neural and adipocyte differentiation, dynamic hydroxmethylation has been associated with lineage-specific distal regulatory regions and represents an early event of enhancer activation (Serandour et al., 2012). Concordantly, a separate study has demonstrated that deletion of Tet2 led to extensive loss of 5hmC and gain of DNA hypermethylation at enhancers and modulates enhancer activity of differentiation-related genes (Ito et al., 2011). However, the roles of TET proteins in FOXA1 recruitment and regulation of prostate lineage-specific enhancers are yet to be delineated.

Here, we show that TET1 is a direct target of FOXA1-mediated transcriptional activation. Further, TET1 physically interacts with the FOXA1 protein and modulates local DNA demethylation that in turn facilitates and stabilizes the recruitment of FOXA1. FOXA1 and TET1 thus form a feed-forward loop that activates lineage-specific enhancers. Not only does this mechanism provide a new perspective on the dynamic functional significance of the newly discovered TET1 DNA hydroxylase, but offer insight into the molecular details underlying FOXA1's ability to fine-tune and modulate lineage-specific enhancer activation. As FOXA1 is a critical regulator and a top mutated gene in multiple cancers such as breast and prostate cancers(Robinson et al., 2013), our study thus forms the framework for future understanding of the roles of TET1 in lineage-specific gene expression and cancer progression.

RESULTS

FOXA1 expression contributes to lineage-specific enhancer activation

To determine the correlation between FOXA1 and active enhancer marks, we re-analyzed previously published FOXA1 (GSE37345), H3K4me2, and H3K27ac ChIP-seq data (GSE27823) (Jin et al., 2014b; Wang et al., 2011) and confirmed that FOXA1 binding sites (FXBS) are indeed enriched for H3K4me2 and H3K27ac (**Figure 1A**). Further, we performed chemical labeling of 5mC and 5hmC followed by deep sequencing (Song et al., 2011; Song et al., 2012b; Zhang et al., 2013), namely TAmC-seq and hMe-Seal-seq, respectively, to map their genomic landscapes in LNCaP cells which express FOXA1. Bioinformatic analysis revealed that FXBS are depleted of 5mC, but enriched for 5hmC, being consistent with previous reports (Serandour et al., 2011). In addition, we found that this correlation was much weaker in two other prostate cell lines namely PrEC and PC-3M, wherein FOXA1 expression is low, suggesting that FOXA1 expression and occupancy might contribute to DNA demethylation at local chromatin (**Figure S1A-C**).

To further elaborate on this, we depleted FOXA1 in LNCaP cells through lentiviral shRNA transduction (**Figure S1D**) and performed chemical labeling of 5mC and 5hmC for subsequent pull down and deep sequencing. Interestingly, on the global scale, there was an approximate 52.3% gain on the number of total 5mC-enriched peaks following FOXA1 knockdown. Moreover, the average intensity of 5mC around all FOXA1-occupied sites also increased upon FOXA1 depletion (**Figure 1B**). Concordantly, active enhancer marks H3K4me2 and H3K27ac were decreased around FXBS following FOXA1 knockdown, supporting reduced enhancer activities (**Figure S1E-F**). We also examined the average intensity of 5hmC around all FXBS but did not observe an obvious difference, which may be due to 5hmC being an intermediate

methylation mark that can be further oxidized to 5fC and 5caC. However, when we performed hierarchical clustering of FXBS based on the epigenetic patterns we found that 5hmC enrichment was indeed decreased in the group with the strongest FOXA1 binding (**Figure 1C**). It can be inferred from these results that FOXA1 may be functioning to alter DNA methylation specifically at regions where it occupies to achieve a de-methylated state, thus potentiating enhancer activation.

FOXA1 positively regulates TET1 gene expression.

As DNA demethylation has recently been shown catalyzed by the TET proteins, we next examined whether TET gene expression is associated with FOXA1. We first performed qRT-PCR analysis of FOXA1 and TET1 transcript across a panel of 12 prostate cell lines (**Figure 2A-B**). Interestingly, like FOXA1, TET1 is in general expressed at much higher levels in AR-positive prostate cancer cell lines such as C4-2B and VCaP cells than in AR-negative cells including DU145 and RWPE. Further analysis showed that TET1 expression level is highly correlated ($r = 0.96$, $P<0.001$) with that of FOXA1 (**Figure S2**). As the correlation between FOXA1 and other TET proteins is relatively weaker, we decided to focus on TET1 in this study.

Since TET1 exhibited a similar expression pattern as FOXA1, we asked whether FOXA1 regulates TET1 gene expression. To test this, we first examined TET1 level in LNCaP cells with control or FOXA1 knockdown. Importantly, both TET1 transcript and protein levels were markedly decreased in LNCaP cells following FOXA1 knockdown (**Figure 2C-D**). Concordantly, depletion of FOXA1 in another independent prostate cancer cell lines C4-2B also resulted in a decrease in TET1 expression **(Figure 2E**). On the other hand, when FOXA1 was overexpressed in 22Rv1 cells through adenovirus infection, TET1 expression was augmented

(**Figure 2F**), which was further validated in another prostate cancer cell line DU145 that contained low endogenous FOXA1 level (**Figure 2G**). To visualize the inductive effect of FOXA1 on TET1 at the cellular level, we performed immunofluorescence staining. TET1 was barely detectable in control DU145 cells infected with empty vector adenovirus (**Figure 2H**, top panel). However, upon infection with adenoviral FOXA1 (Flag-tagged, shown in red), TET1 staining (shown in green) was significantly enhanced (middle panel). Specifically, TET1 was stained positively in the majority of cells that had FOXA1 infection and overexpression, but not in the uninfected cells, as further illustrated in the zoomed-in microscopy images (**Figure 2H**, bottom panel). Taken together, our data support that FOXA1 positively regulates TET1 gene transcription.

TET1 is a direct transcriptional target of FOXA1

To determine how FOXA1 transcriptionally controls TET1 expression, we examined FOXA1 ChIP-seq data previously obtained from LNCaP cells (Tahiliani et al., 2009), and observed a strong FOXA1 binding event within the intragenic region, between exons 3 and 4, of the TET1 gene (**Figure 3A**). Being consistent with FOXA1 as an enhancer regulator that modulates target genes through enhancer-promoter looping, we also found a weak FOXA1 binding event at the TET1 promoter. To validate the results of ChIP-seq, we performed ChIP-qPCR in LNCaP cells and found that FOXA1 is enriched at the TET1 enhancer for nearly 170 fold relative to IgG control, an enrichment level comparable to that at the PSA enhancer, and for about 10 fold at the TET 1 promoter (**Figure 3B**). A similarly strong enrichment of FOXA1 at the TET1 enhancer and promoter was also observed in an additional FOXA1-expressing cell line C4-2B (**Figure S3**). Moreover, upon lentiviral knockdown, FOXA1 binding to its target site for the PSA gene

was greatly diminished as expected, and similarly for TET1 enhancer and promoter, confirming that the ChIP enrichment signal was specific for FOXA1 (**Figure 3C**). Next, to examine whether FOXA1 occupancy at the TET1 enhancer and promoter leads to regulation of their transcriptional activities, we cloned these regions into reporter constructs. Luciferase assays showed that FOXA1 overexpression indeed significantly increased, whereas FOXA1 knockdown decreased, TET1 enhancer and promoter activities (**Figure 3D-E**). To further demonstrate that this regulation is due to FOXA1 occupancy at the TET1 enhancer and promoter, we analyzed the DNA sequences around the FOXA1 binding peaks for FKHD motifs within the TET1 enhancer as well as promoter. Through mutagenesis assays, we generated TET1 enhancer and promoter constructs with mutations to highly conserved FKHD motifs (**Figure 3A, bottom panels**). Importantly, luciferase assays revealed that mutations to the FKHD motifs abolished FOXA1 regulation of TET1 enhancer as well as promoter activities (**Figure 3F**). Taken together, our data support that FOXA1 directly binds to the regulatory elements of TET1 gene to induce its transcription. As FOXA1 contributes to local DNA demethylation (**Figure 1**) and TET1 is a known DNA demethylase, we hypothesized that TET1 may be attributable for DNA demethylation around the FOXA1 binding sites. To test this hypothesis, we started out by examining potential interactions between the FOXA1 and TET1 proteins.

FOXA1 and TET1 proteins physically interact

By use of overexpression systems in 293T cells, we conducted co-immunoprecipitation (co-IP) experiments to assess whether physical interaction is present between ectopic FOXA1 and TET1 proteins. The 293T cells were co-transfected with Flag-tagged TET1 along with FOXA1 or empty vector. Successful expression of the ectopic proteins was confirmed by western blot

analysis of the input lysate. IP using an anti-FOXA1 antibody followed by immunoblotting confirmed successful pulldown of FOXA1 itself as well as the TET1 protein, the latter only in the cells expressing both TET1 and FOXA1 (**Figure 4A**). To demonstrate the interaction through reversal co-IP, we cloned TET1 into the SFB-tagged expression vector, which enabled pulldown of the TET1 protein using S-protein agarose beads and detection by anti-Flag antibodies (Fong et al., 2013). Either SFB-vector control or SFB-TET1 was co-transfected with FOXA1 into 293T cells and their expression was confirmed by western blot analysis of the input lysate. S-protein pulldown followed by western blotting using anti-Flag validated successful enrichment of SFB-tag only or SFB-TET1 (of different sizes) in the corresponding lysates, while immunoblotting using anti-FOXA1 revealed FOXA1 pulldown only in the SFB-TET1-expressing cells (**Figure 4B**), supporting physical interaction between ectopic FOXA1 and TET1 proteins.

Next, we attempted to confirm this interaction between endogenous FOXA1 and TET1 proteins. LNCaP cell nuclear lysate was subjected to IP using rabbit anti-TET1, anti-FOXA1, and IgG control followed by western blotting with mouse anti-TET1 or anti-FOXA1 antibodies. Our results demonstrated that TET1 and FOXA1 antibodies are able to pull down each other, supporting strong protein interactions (**Figure 4C**). To address the potential involvement of DNA in mediating this interaction, we performed co-IP in the presence or absence of ethidium bromide. Notably, our results demonstrated persistent interaction between FOXA1 and TET1 proteins in the presence of ethidium bromide, thus indicating that DNA was not required for their association (**Figure S4**).

To further determine which domains of the TET1 protein are important for its interaction with FOXA1, we generated four Myc-tagged TET1 domain constructs, namely the N-terminal, CXXC, middle and CD domains, which were co-transfected with SFB-tagged FOXA1 into 293T

cells. S-protein pulldown followed by western blot analysis showed that only the TET1 fragment containing the CXXC module was able to bind FOXA1 (**Figure 4D**). On the other hand, we attempted to map out the FOXA1 domain that is responsible for its interaction with the TET1 protein. Similarly, we created three Flag-tagged FOXA1 domain constructs, namely N-terminal, Forkhead (FH), and C-terminal domains, which were co-transfected with SFB-tagged TET1-CXXC domain into 293T cells. Western blot analysis confirmed the expression of various FOXA1 domains of different sizes as expected (**Figure 4E**). S-protein pull down of TET1 followed by western blotting revealed that only the FH-containing domain of FOXA1 protein is able to interact with the TET1-CXXC domain. As the CXXC zinc finger module in Tet3 protein has been shown critical for specific chromatin targeting, while its enzymatic domain modulates its biological function (Xu et al., 2012), we hypothesized that TET1 interaction with FOXA1 through its CXXC domain may be important for its recruitment to FOXA1 binding sites where it carries out hydroxylation on methylated CpG's closeby. Therefore, we next asked whether TET1 regulates DNA demethylation and alters epigenetic modifications around FOXA1 binding sites.

TET1 mediates active epigenetic modification at FOXA1-dependent enhancers

To determine whether TET1 affects the epigenetic environment at FOXA1-occupied enhancers, we first tested whether TET1 is able to co-occupy FOXA1-bound genomic regions. As human anti-TET1 antibody has not been well-established for ChIP, we transfected HA-tagged TET1 into LNCaP prostate cancer cells and performed ChIP using ChIP-grade anti-HA antibody. ChIP-qPCR confirmed much stronger HA (TET1) enrichment at FOXA1 binding sites in cells expressing HA-TET1 than cells transfected with HA-control vector (**Figure 5A**). Next, to examine how TET1 alters DNA methylation around these FOXA1-bound regions, we performed

TET1 knockdown using shRNA (**Figure 5B**). MeDIP-qPCR demonstrated greatly increased DNA methylation around FOXA1 binding sites following TET1 depletion, suggesting that TET1 is critical for the maintenance of the demethylated state of these enhancers (**Figure 5C**). Being concordant with the increase in DNA methylation, the active enhancer marks, H3K4me2 and H3K27ac, were both reduced following TET1 depletion (**Figure 5D-E**). To investigate epigenetic regulation by TET1 on a genome-wide scale, we performed MeDIP-seq and ChIP-seq to globally assay DNA methylation and H3K4me2 enrichment, respectively. Interestingly, next-generation sequencing of enriched DNA fragments demonstrated that TET1 knockdown indeed increased global DNA methylation, while decreased H3K4me2 levels (**Figure 5F**). As TET1 is a DNA demethylase that catalyzes 5mC to 5hmC, we next sought to determine the level of 5hmC in TET1-knockdown cells. Dot blot experiment confirmed significant reduction of total 5hmC abundance in shTET1 cells (**Figure S5**). Chemical labeling followed by pulldown and sequencing (hMe-Seal-seq) revealed a remarkable decrease of total 5hmC-erniched regions upon TET1 knockdown (**Figure 5G**). Further, focused analysis of these epigenetic modifications around FOXA1 binding sites specifically demonstrated an overall increase of DNA methylation but a decrease of 5hmC and H3K4me2 levels (**Figure 5H**). Taken together, our data support that TET1 expression contributes to the activation of FOXA1-target enhancers through mediating DNA de-methylation.

TET1 expression is required for FOXA1 recruitment to target enhancers

Since it has been reported that DNA methylation and removal of H3K4me2 could impair FOXA1 binding (Lupien et al., 2008; Serandour et al., 2011), the changes in DNA methylation and histone modification events observed following TET1 depletion were suggestive of

disrupted FOXA1 recruitment to these regions. To test this, we performed FOXA1 ChIP-seq in control and TET1-knockdown LNCaP cells to determine whether TET1 depletion is able to regulate FOXA1 chromatin targeting. A global assessment of the total binding events before and after TET1 knockdown demonstrated that a significant proportion of FOXA1 binding events were lost upon TET1 depletion (**Figure 6A**). The total number of FXBS was decreased from 76,000 to 55,000. In addition, the average intensity of FOXA1 binding events appeared to be much weaker even for the sites that were not fully abolished (i.e. shared sites) following TET1 knockdown (**Figure 6B**). Genome browser view of several FOXA1-dependent enhancers further illustrated significant loss of FOXA1 occupancy in TET1-depleted cells (**Figure 6C-D & S6A-B**). Meanwhile, DNA methylation at these enhancers was increased as indicated by enhanced 5mC but reduced 5hmC signals, while active enhancer mark H3K4me2 was decreased, being concordant with the genome-wide switch to repressive chromatin state as shown in Figure 5F-G. Moreover, ChIP-qPCR confirmed that TET1 knockdown significantly decreased FOXA1 occupancy at multiple target enhancers (**Figure 6E**).

As TET1 interacts with the FOXA1 protein through its CXXC domain but is known to carry out enzymatic activities through its CD domain, we next attempted to understand mechanistically whether CD-mediated DNA demethylation is sufficient to facilitate FOXA1 recruitment to target enhancers. A recent study has reported an interesting and important observation that the CD domain of TET proteins induces massive global DNA demethylation, whereas the function of full-length TET1 is much restricted to unmethylated CpG islands (Jin et al., 2014a). We thus predict that CD domain may be able to restore FOXA1 recruitment in TET1-knockdown cells. To test this, we overexpressed the CD domain in LNCaP cells with TET1 knockdown. ChIP-qPCR confirmed that FOXA1 binding at target enhancers was decreased by TET1 knockdown,

which, importantly, can be fully rescued by concomitant CD domain overexpression (**Figure 6F**). Taken together, our data support that TET1 facilitates FOXA1 recruitment to target enhancers through active demethylation.

DISCUSSION

FOXA1 is a critical regulator of hormone-mediated gene expression in prostate and breast cancers. Much efforts have been devoted to understand the molecular basis for FOXA1's activity as a pioneer factor, and studies in the past two decades have helped to paint a clearer picture of how FOXA1 activity is dependent on a number of epigenetic signatures that exhibit lineage specificity (Lupien et al., 2008). Although FOXA1 has been shown to impose some effects on the epigenetic signatures around target enhancers (Lupien et al., 2008), the molecular mechanisms by which FOXA1 alters the epigenetic switch remains largely unknown. In the present study, we show that FOXA1 is able to directly regulate the transcription of *TET1* gene. Further, FOXA1 physically interacts with the TET1 protein, leading to DNA demethylation and H3K4me2/H3K27ac modifications at FOXA1-target enhancers. These changes in the epigenetic environment on the other hand enhance FOXA1 recruitment. Therefore, our data support a model wherein FOXA1, a reader of active epigenetic marks on enhancers, in turn contributes to enhancer activation, thus forming a positive feedback loop (**Figure 7**). We also showed a feed-forward loop between FOXA1 and TET1 where FOXA1 induces TET1 expression and binding at lineage-specific enhancers, which in turn facilitates and stabilizes FOXA1 recruitment. Accompanying changes in DNA methylation are reductions in H3K4me2 and H3K27ac upon FOXA1 depletion. Whether these are secondary to DNA demethylation or FOXA1/TET1 may regulate histone methyltransferases such as MLL are areas for future investigation.

TET1 has been recently implicated in regulating enhancer activation and lineage differentiation through DNA demethylation (Hon et al., 2014; Ito et al., 2011), the underlying mechanism of which, however, remains elusive. In this report, using prostate cancer cells as a model system, we demonstrated that TET1 contributes to FOXA1 recruitment to prostate-specific enhancers by modulating local epigenetic switch. In future studies, it will be interesting to investigate and compare how TET1 regulates epigenetic marks and FOXA1 recruitment in breast cancer, since FOXA1 has been shown to bind distinct, lineage-specific enhancers in prostate and breast cells (Lupien et al., 2008). In addition, this study will pave the way to further investigation of how TET1, through modulation of FOXA1-dependent enhancer activation, regulates hormone-dependent gene expression and prostate and breast cancer progression.

The CXXC domain of TET proteins has been shown to be critical for specific chromatin targeting, while the CD domain modulates its enzymatic activity (Xu et al., 2012). Further, a recent study has reported that the full-length TET1 protein preferably binds to unmethylated CpG islands through its CXXC domain (Jin et al., 2014a). Being consistent with these reports, we found that FOXA1 interacts with TET1 protein through its CXXC domain. Such interaction may be critical for targeting TET1 to prostate-specific enhancers denoted by FOXA1, which may be intereting lines for further investigation utilizing various TET1 deletion constructs and ChIP-seq experiments. Moreover, TET1 might similarly interact with other lineage-defining transcription factors and get recruited to distinct, lineage-specific enhancers in different cell types. By contrast, overexpression of the TET1 CD domain alone has been shown to induce masive global DNA demethylation (Jin et al., 2014a). Indeed, in our study we found overexpression of CD domain is able to rescue the effects of TET1 knockdown on FOXA1 recruitment to target enhancers. Therefore, through interaction with other transcription factors,

TET1 achieves its specificity to bind selected enhancers, where it carries out its role in the maintenance of hypomethylated landscape and regulation of lineage differentiation.

In conclusion, FOXA1 is a multipotent pioneer transcription factor, which is impressively capable of chromatin remodeling through not only histone displacement but also DNA demethylation by employing the DNA hydroxylase TET1. Collectively, through regulation of TET1 expression and function, FOXA1 is able to control the epigenetic signatures present at its cis-regulatory elements through a feed-forward loop, ultimately giving rise to chromatin relaxation and enhancer activation.

EXPERIMENTAL PROCEDURES

Cell lines, Plasmids and Antibodies

Prostate cancer cell lines LNCaP, VCaP, 22Rv1, BPH1, RWPE-1, DU145, and human embryonic kidney cell line HEK293T cells were obtained from American Type Culture Collection (ATCC) and cultured in either RPMI1640 or DMEM with 10% FBS. For FOXA1 and TET1 FL and domain constructs, human FOXA1 and TET1 cDNA were amplified by reverse transcription PCR from LNCaP cells and pENTR223 TET1 (Harvard Plasmid), respectively, and cloned into the entry vector pCR8/GW/TOPO (Invitrogen). Adenoviral construct expressing FOXA1 was generated by recombining pCR8-FOXA1 with pAD/CMV/V5 using LR Clonase II (Invitrogen). Overexpression constructs for TET1 were generated by recombination of pCR8-TET1 with NTSFB destination vector or pLenti CMV/TO Puro DEST (Addgene plasmid 17293). The pGIPZ lentiviral control and FOXA1 shRNAs were purchased from Open Biosystems. Sequences for scramble (5'-GCGCGCTTTGTAGGATTCG-3') and TET1 (5'-

GTGGAGAAGTGGACACAAA-3') shRNA were kindly provided by Dr. Debabrata Chakravarti (Northwestern University), and cloned into pLKO lentiviral vector.

The antibodies used in this study include anti-FOXA1 (ab23738) and anti-GAPDH (ab9385) from Abcam, anti-TET1 (GTX627420 and GTX124207) from GeneTex, anti-FLAG (F1804 and F7425) from Sigma, anti-c-Myc (sc-789x) from Santa Cruz, anti-HA (ab9110) from Abcam, anti-alpha Tubulin (sc-32293) from Santa Cruz, anti-5mC (BI-MECY-0100) from Eurogentec, anti-5hmC (39769) from Active Motif, anti-H3K4me2 (07-030) from Millipore, anti-H3K27ac (ab4279) from Abcam.

Luciferase Reporter Assay

TET1 promoter and enhancer luciferase reporter assays were conducted according to the manual of Dual-Luciferase Reporter Assay System from Promega. Briefly, LNCaP cells were seeded in a 24-well plate and co-transfected with the Renilla expression plasmid pRL-TK and the reporter constructs for TET1 promoter and enhancer in pGL4 vector. Cells were infected with LacZ (control) or FOXA1 adenovirus for 48 hours to assess the effect of FOXA1 overexpression on luciferase activity. Conversely, to look at FOXA1 depletion effect, lentiviral-transduced shCtrl and shFOXA1 LNCaP cells were used for co-transfection of reporter constructs. Luciferase activities were determined 48-72 hours post-transfection and normalized against Renilla internal control values.

Immunofluorescence Staining

Cells were fixed with 4% formaldehyde for 15min at RT and then permeabilized in 0.1% Triton X-100 for 15min at RT. Cells were then washed by PBS for three times, followed by incubation with 5% normal goat serum for 30min at RT. Subsequently, cells were incubated with primary

antibody, the anti-FLAG antibody (Sigma) and anti-TET1 (Genetex), for 2h at RT. After washing three times with PBS, cells were incubated with secondary antibody, Alexa Fluor 488 and 594 goat anti-rabbit IgG (Invitrogen), for 1h at RT. Finally, cells were washed three times with PBS and mounted using Prolong Gold Antifade Reagent (Invitrogen).

Co-Immunoprecipitation

Nuclear proteins were extracted from 293T or LNCaP cells (details provided in Supplementary Methods). For S protein pulldown, nuclear extracts were incubated with 30ul S-protein agarose beads (Millipore) for 3hr at 4^{o}C. The beads/protein complex was then washed 4 times, and eluted with 30ul 2X SDS sample buffer and subjected to western blot analysis. For LNCaP endogenous Co-IP, nuclear extracts were incubated with 2ug antibodies, anti-FOXA1 (Abcam), anti-TET1 (Genetex) and anti-rabbit IgG (Santa Cruz) overnight at 4^{o}C. Dynabeads Protein A (Life Technologies), 25ul per IP, were added the next day and incubated for 1hr at 4^{o}C. Similarly, the beads/protein complex was washed 4 times, and eluted with 30ul 2X SDS sample buffer and subjected to western blot analysis.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments were carried out as previously described. All primers (listed in Supplementary Information) were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/), synthesized by Integrated DNA Technologies, and used for SYBR Green based real-time PCR. ChIP-quantitative PCR enrichment of target loci was normalized to input DNA and reported as % input ±SEM. ChIP DNA was prepared into libraries according to standard protocols using Bioo Scientific's DNA Sample Kit (cat. no. 514101). Libraries were sequenced using Illumina Hi-Seq

platforms. Sequence reads were aligned to the Human Reference Genome (assembly hg19) using Burrows − Wheeler alignment tool (bwa) version 0.6.1.

Methylated DNA Immunoprecipitation (MeDIP)

MeDIP was performed as previously described (Yu et al., 2010). Total genomic DNA was extracted using QIAamp DNA Mini Kit (Qiagen), and sonicated to obtain fragments between 300-1000bp. Dynabeads M-280 Sheep anti-Mouse IgG (Invitrogen) were incubated with an anti-5-methylcytidine antibody (BI-MECY_0100, Eurogentec, Fremont, CA, USA) overnight at 4^{o}C. The following day, 4ug of sheared DNA was denatured by boiling at 95 ^{o}C for 10min followed by rapid cooling on ice, and subsequently added to the beads/antibody complex. On day 3, the beads were washed 3 times with PBS+0.05% Triton X-100 and eluted from beads by incubation at 65^{o}C for 5min in 150ul elution buffer (TE+1% SDS). Elution was repeated for a total of two times. Total eluates were treated with proteinase K and incubated at 50^{o}C for 2hr. QIAquick PCR purification kit (Qiagen) was used to purify the eluted DNA, and lastly qPCR was used to determine the enrichment of target genomic regions using gene-specific primers (listed in Supplemental Information). Enrichment of target loci was normalized to input DNA and reported as % input ±SEM.

5mC and 5hmC Chemical Labeling (TAmC and hMe-Seal)

5mC labeling experiments were done as previously published (Zhang et al., 2013), and 5hmC labeling experiments were performed as previously described (Song et al., 2012a). Briefly, genomic DNA was fragmented to an average of 400bp and was incubated with 50mM HEPES buffer (pH 7.9), 25mM MgCl2, 100mM UDP-6-N3-Glc, and 2mM βGT for 1hr at 37°C. The labeled DNA was purified by the QIAquick Nucleotide Removal kit (Qiagen) and eluted in H2O. The click chemistry was performed with the addition of 150mM of disulfide-biotin, and the

mixture was incubated for 2hr at 37°C. The labeled DNA fragments were then purified by the QIAquick Nucleotide Removal kit (Qiagen) and enriched by Dynabeads Streptavidin C1 (Invitrogen), and subsequently released by DTT treatment. The enriched DNA fragments were first purified by Micro Bio-Spin 6 spin columns (Bio-Rad) followed by MinElute PCR Purification Kit (QIAGEN).

AUTHOR CONTRIBUTIONS

Y.A.Y, K.F., J.K., S.L., C.S., B.S., and B.Z. performed the experiments. J.C.Z performed all bioinformatics analysis. J.Y. conceived and supervised the project. J.Y. and Y.A.Y. designed the experiments. C.H. supervised the TAmC-seq and hMe-Seal-seq experiments. J.Y., Y.A.Y, and J.C.Z generated the figures and wrote the manuscript. All authors discussed the results, commented on the manuscript, and declare no conflicts of interest.

ACKNOWLEDGEMENTS

We would like to thank Dr. Debabrata Chakravarti for providing the shTET1 construct. This work was supported in part by the U.S. Department of Defense W81XWH-13-1-0319 (to J.Y.), the Research Scholar Award RSG-12-085-01 (to J.Y.) from the American Cancer Society, and the National Institutes of Health R01HG006827 (C.H.). C.H. is an investigator of the Howard Hughes Medical Institute. Y.A.Y. was supported in part by the NIH/NCI training grant T32 CA09560 and J.K. was supported in part by the NIH Training Program in Oncogenesis and Developmental Biology (T32 CA080621).

REFERENCES

Bernardo, G.M., and Keri, R.A. (2012). FOXA1: a transcription factor with parallel functions in development and cancer. Bioscience reports *32*, 113-130.

Bernardo, G.M., Lozada, K.L., Miedler, J.D., Harburg, G., Hewitt, S.C., Mosley, J.D., Godwin, A.K., Korach, K.S., Visvader, J.E., Kaestner, K.H., *et al.* (2010). FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development *137*, 2045-2054.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schubeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet *9*, e1003994.

Fong, K.W., Leung, J.W., Li, Y., Wang, W., Feng, L., Ma, W., Liu, D., Songyang, Z., and Chen, J. (2013). MTR120/KIAA1383, a novel microtubule-associated protein, promotes microtubule stability and ensures cytokinesis. J Cell Sci *126*, 825-837.

Gao, N., Ishii, K., Mirosevich, J., Kuwajima, S., Oppenheimer, S.R., Roberts, R.L., Jiang, M., Yu, X., Shappell, S.B., Caprioli, R.M., *et al.* (2005). Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. Development *132*, 3431-3443.

Hon, G.C., Song, C.X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.A., Ye, Z., Mao, S.Q., Wang, B.A., *et al.* (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell *56*, 286-297.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M.R., Barton, M.C., and Issa, J.P. (2014a). TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acids Res *42*, 6956-6971.

Jin, H.J., Zhao, J.C., Wu, L., Kim, J., and Yu, J. (2014b). Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nat Commun *5*, 3972.

Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. Nature reviews. Cancer *12*, 381-385.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958-970.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep *12*, 1184-1195.

Robinson, J.L., Holmes, K.A., and Carroll, J.S. (2013). FOXA1 mutations in hormone-dependent cancers. Frontiers in oncology *3*, 20.

Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L., *et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic acids research *40*, 8255-8265.

Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., *et al.* (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res *21*, 555-565.

Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2012a). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nature methods *9*, 75-77.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., *et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature biotechnology *29*, 68-72.

Song, C.X., Yi, C., and He, C. (2012b). Mapping recently identified nucleotide variants in the genome and transcriptome. Nature biotechnology *30*, 1107-1116.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.
Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., *et al.* (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390-394.
Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., *et al.* (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome research *22*, 1798-1812.
Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., *et al.* (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200-1213.
Yu, J., Cao, Q., Yu, J., Wu, L., Dallol, A., Li, J., Chen, G., Grasso, C., Cao, X., Lonigro, R.J., *et al.* (2010). The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. Oncogene *29*, 5370-5380.
Zhang, L., Szulwach, K.E., Hon, G.C., Song, C.X., Park, B., Yu, M., Lu, X., Dai, Q., Wang, X., Street, C.R., *et al.* (2013). Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat Commun *4*, 1517.

FIGURE LEGENDS

Figure 1. FOXA1 contributes to enhancer activation through epigenetic modifications.

(A) Epigenetic signatures of FOXA1 binding sites in control and shFOXA1 LNCaP cells. FOXA1 and H3K4me2/H3K27ac ChIP-seq data were obtained from publicly available datasets GSE37345 and GSE27823, respectively. Genomic landscapes of 5mC and 5hmC were determined by chemical labeling followed by pull down and deep sequencing, as detailed in method for TAmC- and heMe-Seal-seq, respectively. ChIP-seq read intensities of indicated epigenetic marks around (±5 kb) FOXA1 binding sites (FXBS) in control (shCtrl) and FOXA1-knockdown (shFOXA1) cells were together clustered, leading to three major groups as shown by heatmap.

(B, C) Average intensity plots of 5mC (**B**) and 5hmC (**C**) enrichment around the three groups of FXBS shown in **A**. Red indicates control cells and blue FOXA1-knockdown.

Figure 2. FOXA1 induces TET1 gene expression.

(**A, B**) Correlated FOXA1 and TET1 gene expression in prostate cells. RNA was extracted from a panel of 12 prostate cell lines and analyzed by qRT-PCR for FOXA1 (**A**) and TET1 (**B**) gene expression. Data shown is mean ± SEM of technical replicates from one representative experiment out of three.

(**C, D**) TET1 transcript (**C**) and protein (**D**) are down-regulated following FOXA1 knockdown in LNCaP cells. LNCaP cells were infected with shCtrl or shFOXA1 lentivirus and subsequently subjected to qRT-PCR and western blot analysis. Data shown is one representative out of triplicate experiments.

(**E**) TET1 is down-regulated by FOXA1 knockdown in C4-2B cells. C4-2B cells were infected with shCtrl or shFOXA1 lentivirus for 8 hours followed by puromycin selection for 4 days, and subsequently subjected to qRT-PCR and western blot analysis. Data shown is one representative out of triplicate experiments.

(**F, G**) TET1 is up-regulated following FOXA1 overexpression. The 22Rv1 (**F**) and DU145 (**G**) cells were infected with LacZ or FOXA1 adenovirus for 48 hours and immunoblot was performed to assess FOXA1 and TET1 protein levels.

(**H**) Positive TET1 staining in FOXA1-expressing cells. DU145 cells were infected with LacZ control (i-iii) or Flag-tagged FOXA1 (iv-vi) adenovirus for 48 hours and then subjected to Immunofluorescence co-staining of FOXA1 (red) and TET1 (green). Bottom panel shows zoomed-in region containing both FOXA1-uninfected and -infected cells.

Figure 3. TET1 is a direct transcriptional target of FOXA1.

(**A**) ChIP-seq showing FOXA1 binding events at TET1 promoter and enhancer. FOXA1 ChIP-seq was conducted in LNCaP cells and FOXA1 binding events were identified by HOMER and visualized in UCSC Genome Browser. FKHD motifs (indicated by red box) near FOXA1 binding sites were determined by JASPAR. DNA fragments containing FOXA1 binding sites at the TET1 promoter (pTET1) and enhancer (eTET1) were each cloned into pGL4 luciferase reporter construct with wildtype (WT) or mutated (mut) FKHD motif (mutated nt shown in red at the bottom panel).

(**B**) ChIP-PCR validation of FOXA1 binding to TET1 enhancer and promoter in LNCaP cells. ChIP was performed using anti-FOXA1 and anti-IgG antibodies in LNCaP cells. ChIP-qPCR was performed using primers flanking the FOXA1 binding peaks at the TET1 enhancer (eTET1)

and promoter (pTET). PSA is used as a positive control while KIAA0066 a negative control. Data shown is mean ± SEM of technical replicates from one representative experiment out of 3.

(**C**) FOXA1 occupancy at TET1 promoter and enhancer was decreased by FOXA1 knockdown. ChIP-qPCR using anti-FOXA1 antibody was carried out in control and FOXA1-depleted LNCaP cells. Data shown is mean ± SEM of technical replicates from one representative experiment out of 3.

(**D, E**) FOXA1 positively regulates TET1 enhancer and promoter activities. TET1 enhancer and promoter reporter constructs were transfected into LNCaP cells with control of FOXA1 overexpression (**D**) and LNCaP cells with control of FOXA1 knockdown (**E**) for 48 hours. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown is mean ± SEM of two independent experiments.

(**F**) FKHD motif is required for FOXA1-induced TET1 promoter and enhancer luciferase activities. Control and FOXA1-overexpressing LNCaP cells were transfected with either wildtype or mutated (depicted in **A**) TET1 promoter and enhancer reporter constructs. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown is mean ± SEM of two independent experiments.

Figure 4. FOXA1 and TET1 proteins physically interact.

(**A**) Immunoprecipitation of ectopic FOXA1 pulled down TET1 protein. The 293T cells were transfected with Flag-TET1, either alone or together with FOXA1, for 48 hours and then subjected to immunoprecipitation using an FOXA1 antibody. Whole cell (Input) and IP-enriched lysates were then analyzed by western blotting using anti-Flag (TET1) and anti-FOXA1 antibodies.

(**B**) Ectopic TET1 immunoprecipitation pulled down FOXA1 protein. The 293T cells were co-transfected with FOXA1 and SFB-tagged empty vector (EV) or TET1 for 48 hours before immunoprecipitation using S beads, which will pull down SFB-EV or SFB-TET1. The input and IP-enriched cell lysates were then subjected to western blotting using anti-FOXA1 and anti-Flag (for SFB-EV or SFB-TET1) antibodies.

(**C**) Endogenous FOXA1 and TET1 proteins interact in LNCaP cells. LNCaP cells were subjected to immunoprecipitation using anti-FOXA1, anti-TET1, and IgG control, followed by western blotting of FOXA1 and TET1 proteins.

(**D**) TET1 CXXC domain interacts with the FOXA1 protein. 293T cells were co-transfected with SFB-FOXA1 along with various Myc-tagged TET1 domain constructs. The expression of TET1 domains in whole cell lysate (input) was confirmed by western blotting using anti-Myc. Cell lysates were then subjected to S pull down (of FOXA1) and subsequently western blot analysis using anti-FOXA1 and anti-Myc antibodies.

(**E**) FOXA1 FH (Forkhead-containing) domain interacts with TET1 CXXC domain. 293T cells were co-transfected with SFB-CXXC along with various Flag-tagged FOXA1 domain constructs and subjected to S pull down (of TET1-CXXC) followed by western blotting using an anti-Flag antibody.

Figure 5. TET1 mediates active epigenetic modification at FOXA1-bound enhancers.

(**A**) TET1 co-occupies FOXA1 binding sites. LNCaP cells were transfected with HA-tagged empty vector or TET1 constructs and were subsequently used for ChIP with anti-HA antibody. HA ChIP-qPCR was performed using primers flanking a number of FOXA1 binding sites. Data shown is mean ± SEM of technical replicates from one representative experiment out of 3.

(**B**) Western blots confirming TET1 knockdown. LNCaP cells were infected with either scramble or shTET1 lentivirus followed by puromycin selection for 4 days before western blot analysis. Tubulin is used as a loading control.

(**C-E**) TET1 knockdown led to altered epigenetic signatures at FXBS. LNCaP cells with control or TET1 knockdown were subjected to MeDIP (**C**) and ChIP using anti-H3K4me2 (**D**) and anti-H3K27ac (**E**) antibodies, followed by qPCR analysis with site-specific primers. Data shown is mean ± SEM of technical replicates from one representative experiment out of 2.

(**F, G**) Venn Diagrams showing alterations in global genomic regions enriched for 5mC, H3K4me2 (**F**) and 5hmC (**G**) following TET1 knockdown. LNCaP cells with control or TET1 knockdown were subjected to MeDIP-seq, H3K4me2 ChIP-seq, and hMe-Seal-Seq for genome-wide location analysis of 5mC, H3K4me2, and 5hmC, respectively, which were subsequently compared between control and TET1-depleted cells.

(**H**) Average intensity plots of normalized MeDIP-seq, H3K4me2-seq and hMe-Seal-seq reads around (±5 kb) FOXA1 binding sites. Red indicates control cells and blue TET1-knockdown.

Figure 6. TET1 is required for FOXA1 recruitment to lineage-specific enhancers.

(**A**) Venn diagram showing overlap of FOXA1 binding sites in control and FOXA1-knockdown LNCaP cells.

(**B**) Average FOXA1 ChIP-seq read intensity around (±500bp) shCtrl-only, shared and shTET1-only FXBS identified from overlap Venn diagram in (**A**). Red indicates control cells and blue TET1-knockdown.

(**C, D**) Genome browser views of epigenetic modifications at the regulatory regions of FOXA1-target genes SNAIL (**C**) and TET1 itself (**D**). MeDIP-seq (5mC), hMe-Seal-seq (5hmC),

H3K4me2 and FOXA1 ChIP-seq were performed in control and TET1-knockdown LNCaP cells. For each mark, the shCtrl and shTET1 tracks are shown on the same scale (Y-axis) for visual comparison of enrichment.

(**E**) TET1 depletion attenuates FOXA1 recruitment to target enhancers. ChIP-qPCR was performed in control and shTET1 LNCaP cells using anti-FOXA1 antibody. Data shown is mean ± SEM of technical replicates from one representative out of triplicate experiments.

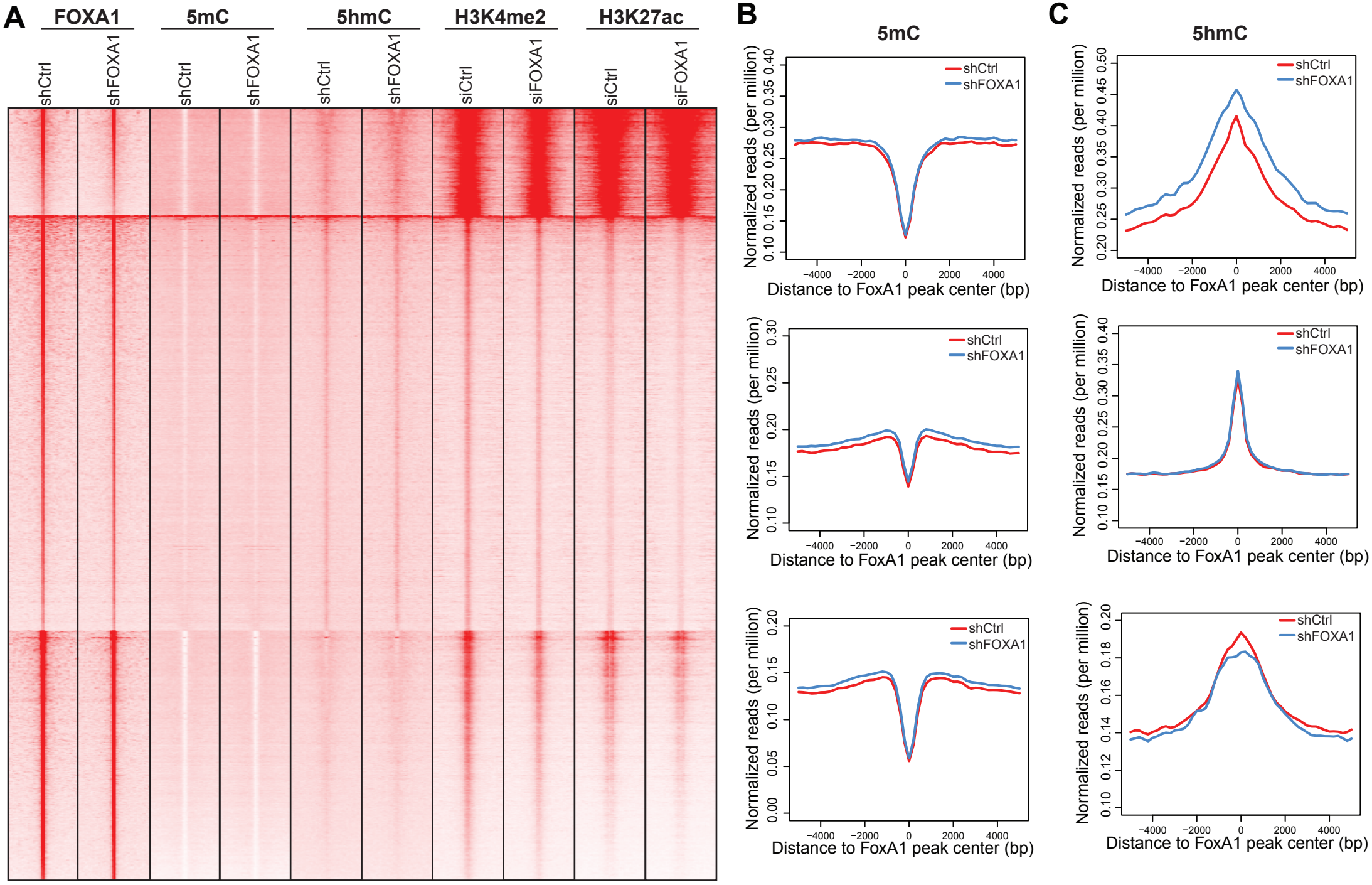
(**F**) Impaired FOXA1 recruitment in TET1-depleted cells is restored by TET1 CD overexpression. LNCaP cells were subjected to control or TET1 knockdown with or without concomitant TET1 CD overexpression. TET1 knockdown and CD domain (Flag-CD) overexpression were confirmed by western blot analysis (inset). Cells were subsequently used for ChIP with an anti-FOXA1 antibody followed by qPCR analysis. Data shown is mean ± SEM of technical replicates from one representative of duplicate experiments.

Figure 7. Schematic model depicting feed-forward regulation between FOXA1 and TET1 in lineage-specific enhancer activation.

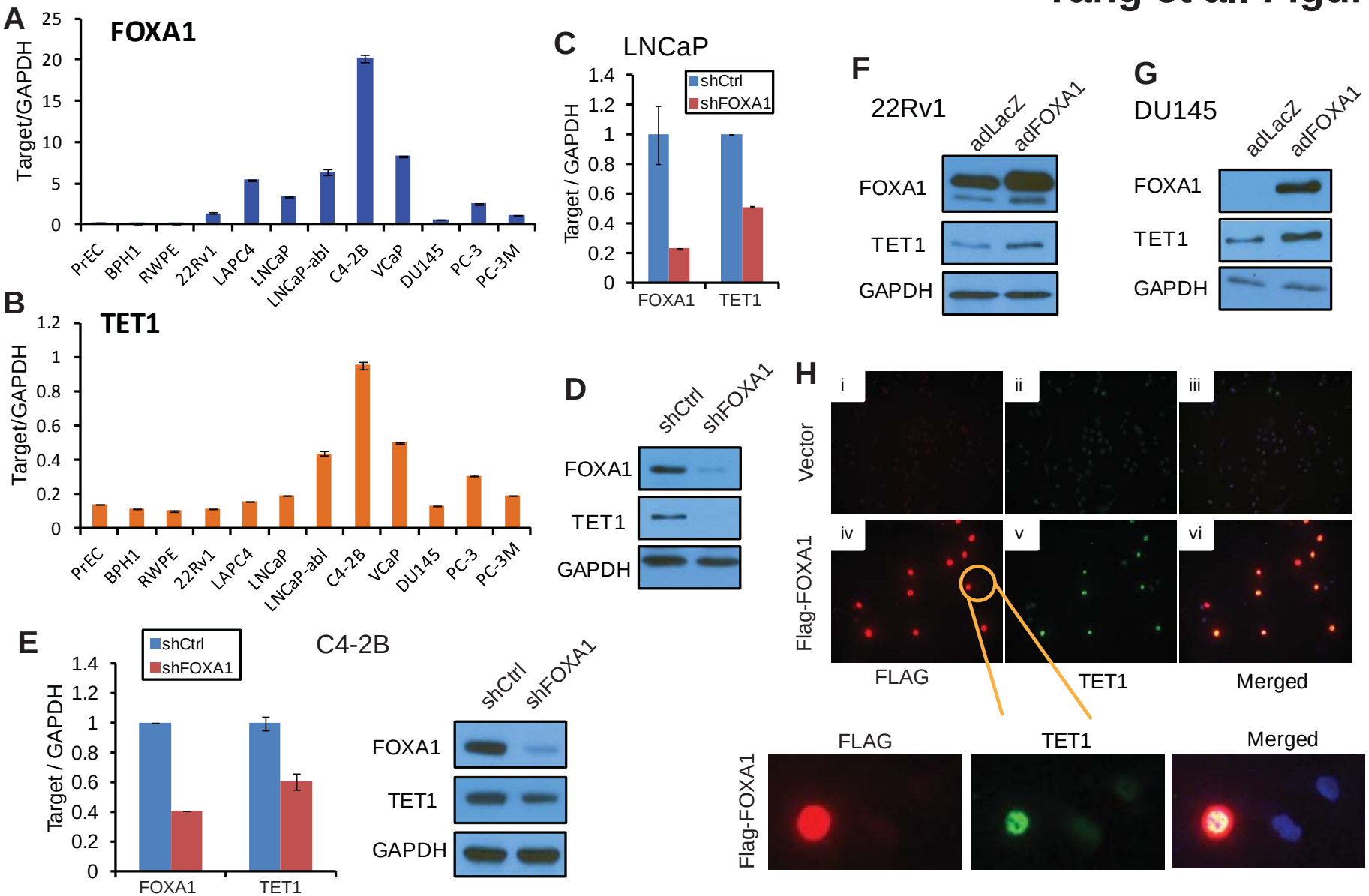
FOXA1 protein occupies at an intragenic enhancer of the TET1 gene to directly induce TET1 expression. Through direct interaction with FOXA1 protein, TET1 modulates DNA demethylation and subsequently H3K4 methylation and H3K27 acetylation at FOXA1-target enhancers, which in turn facilitates FOXA1 recruitment. Thus, FOXA1 and TET1 form a positive feedback loop in lineage-specific enhancer activation. FOXA1 is not only a reader, but also a writer of epigenetic signatures at lineage-specific enhancers.

Figure

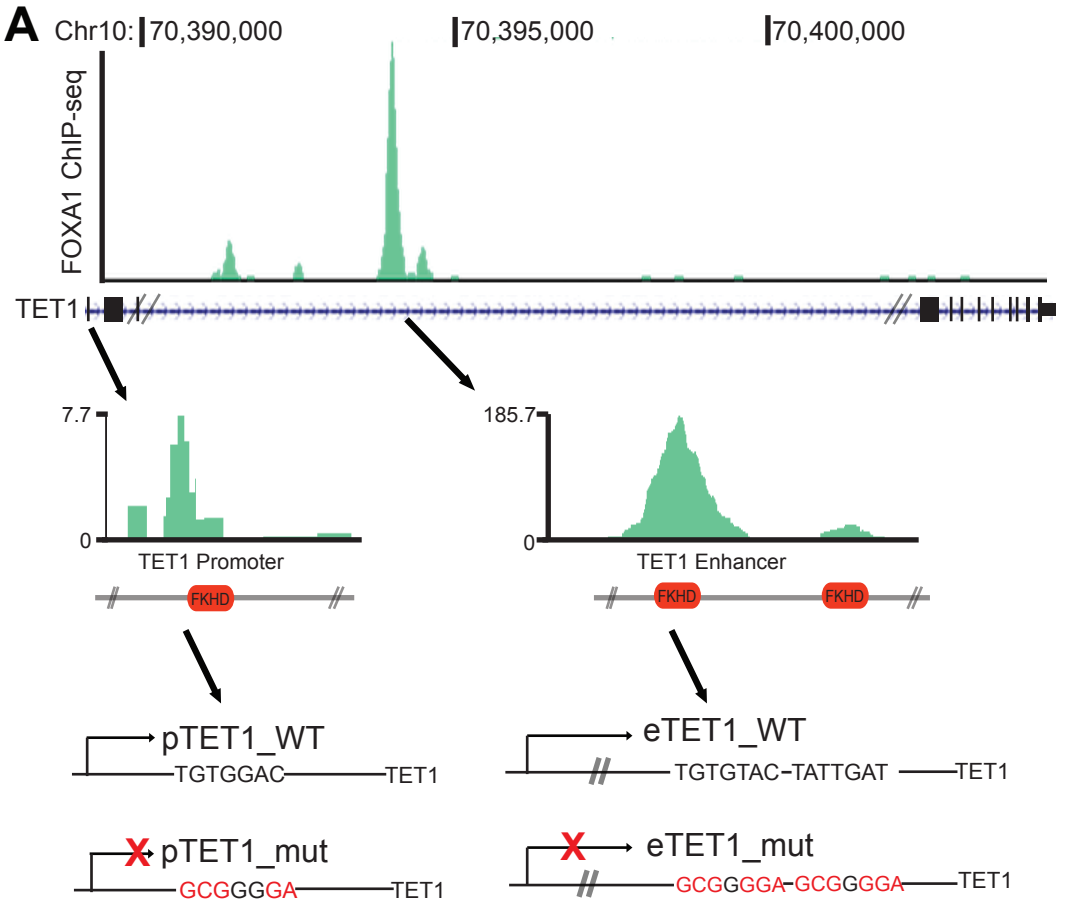
Yang et al. Figure 1

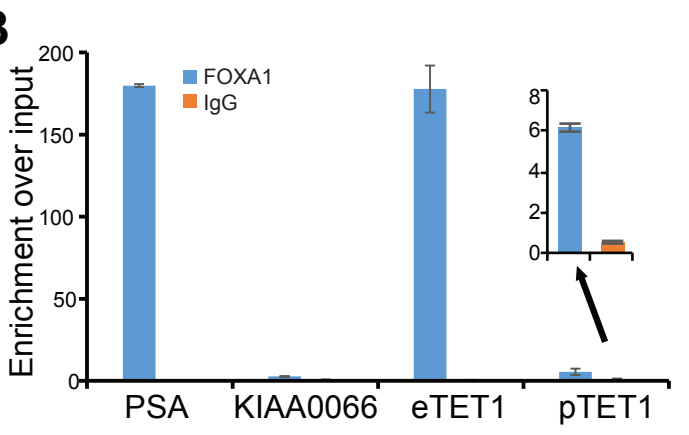


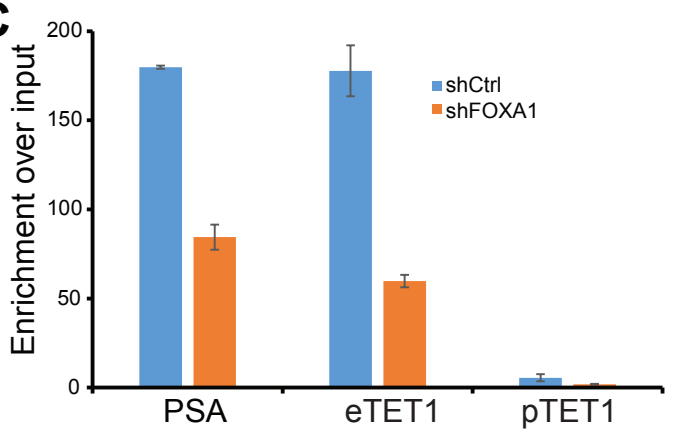
Yang et al. Figure 2

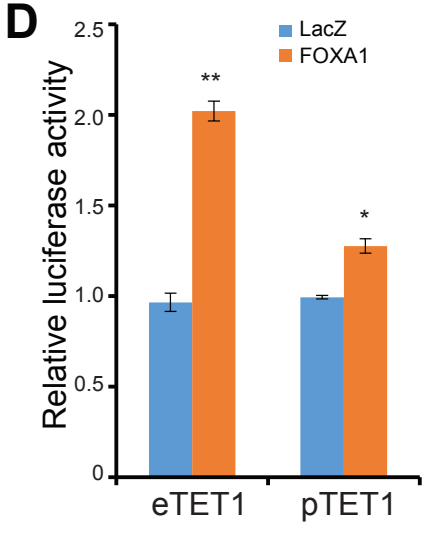


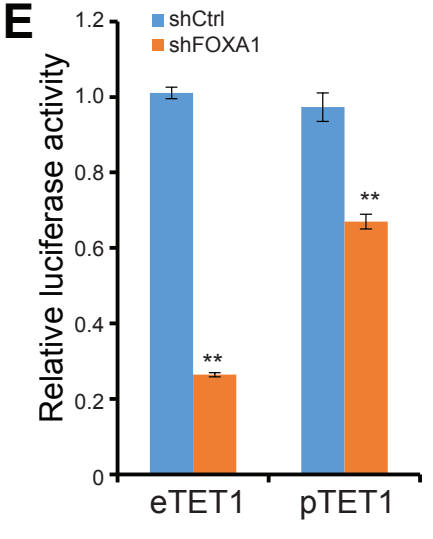
Yang et al. Figure 3

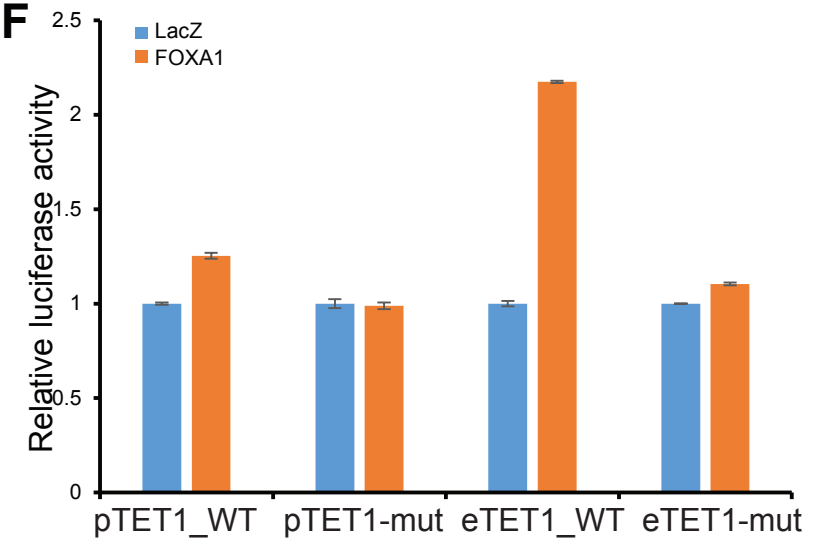




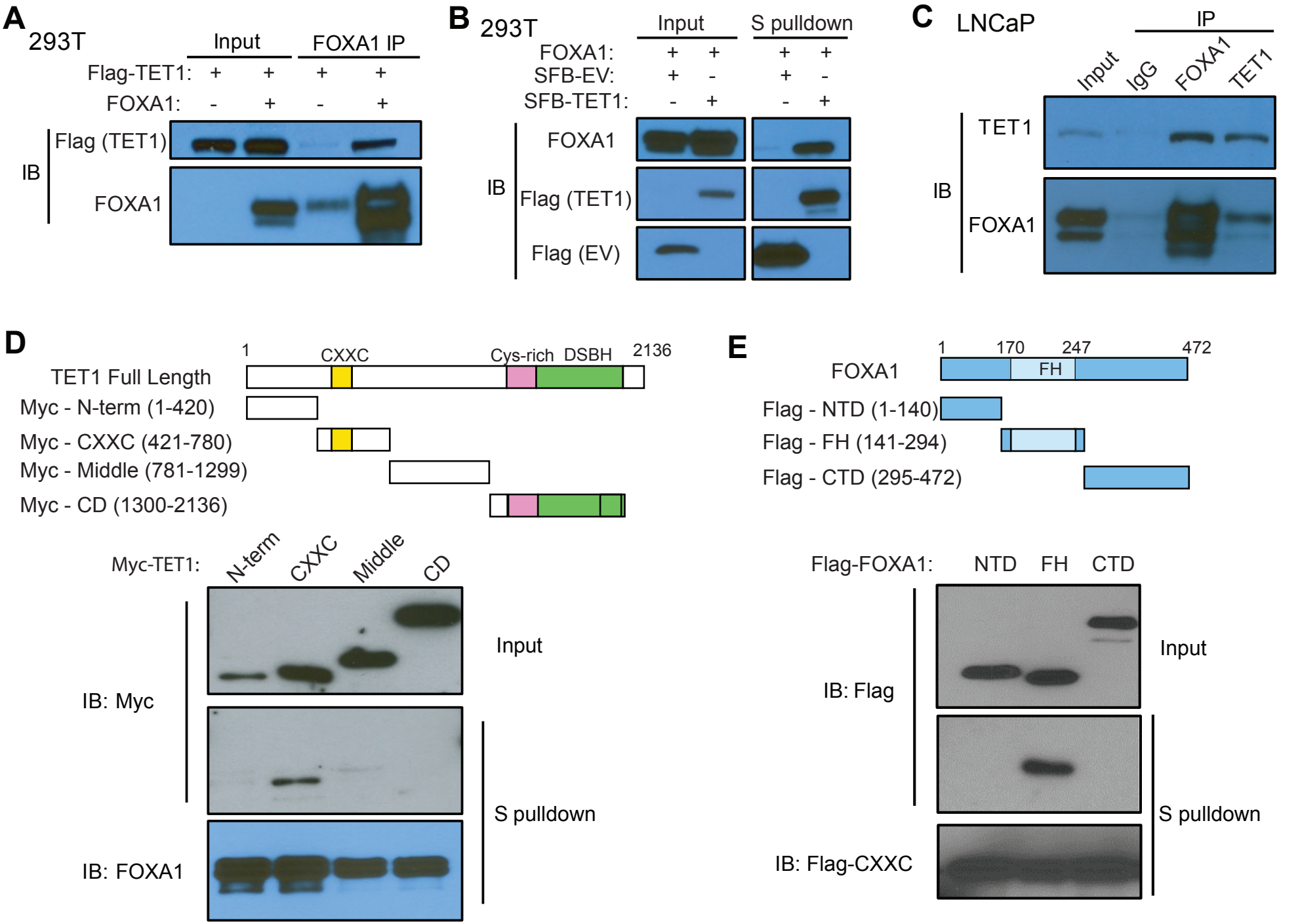




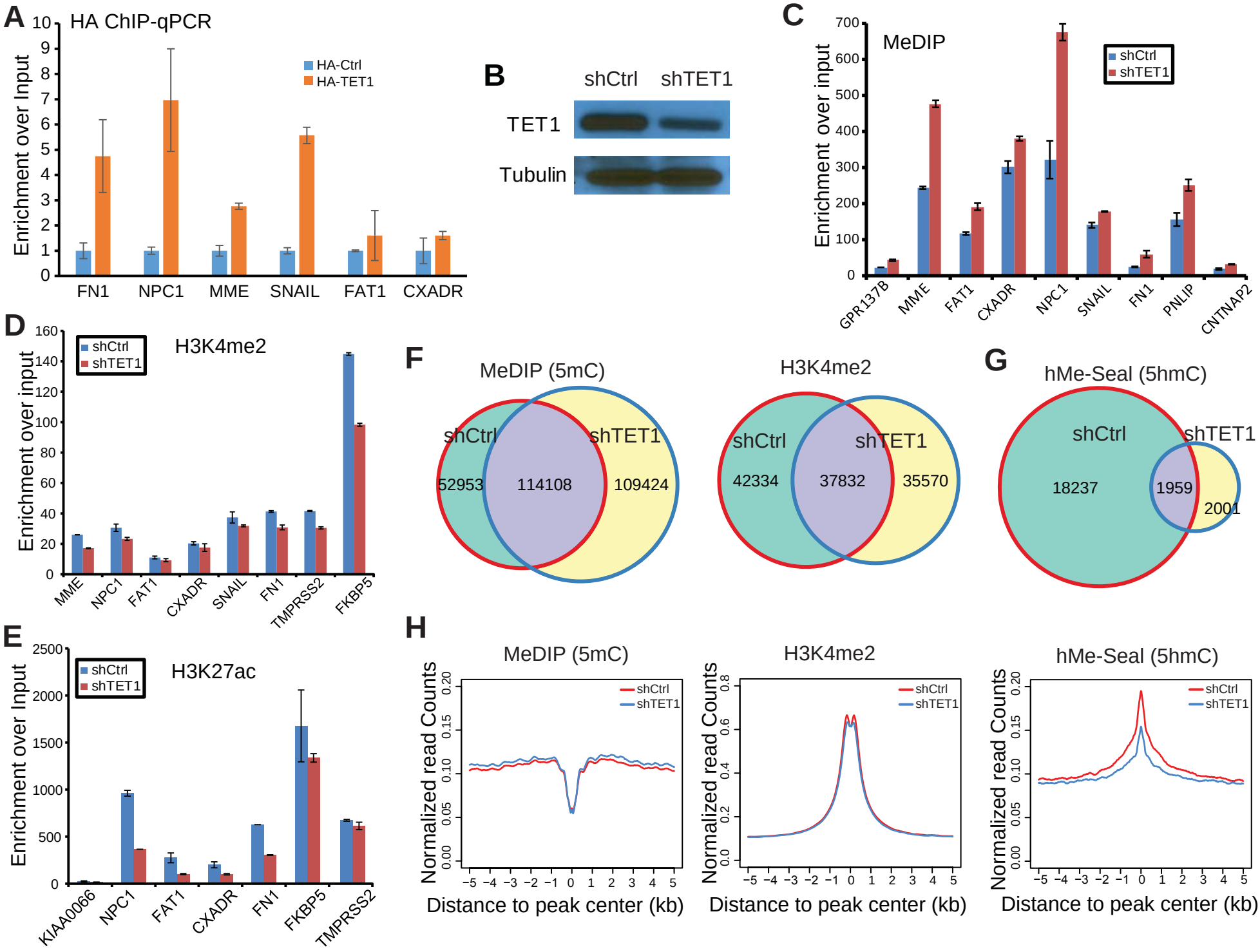




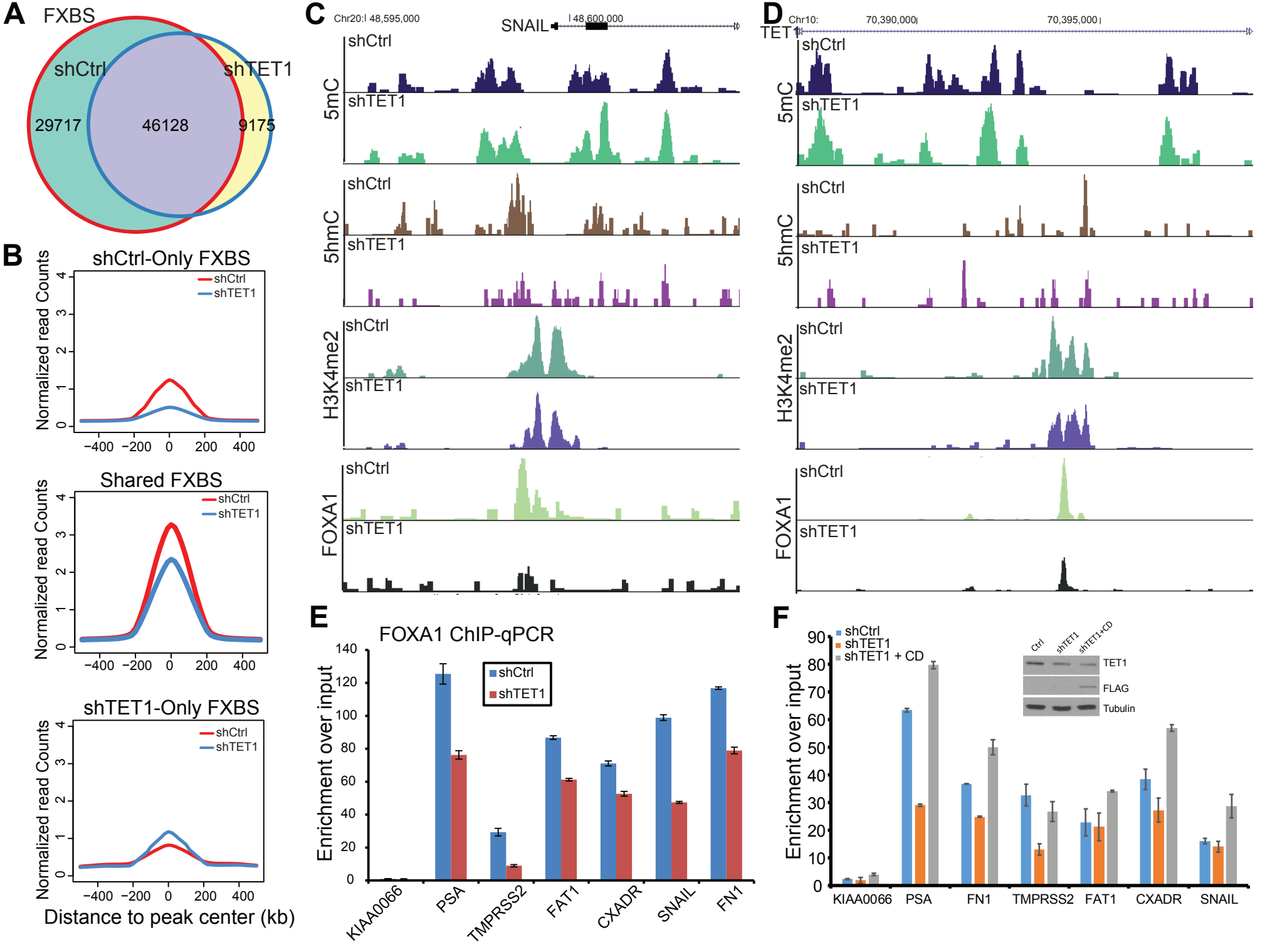
Yang et al. Figure 4



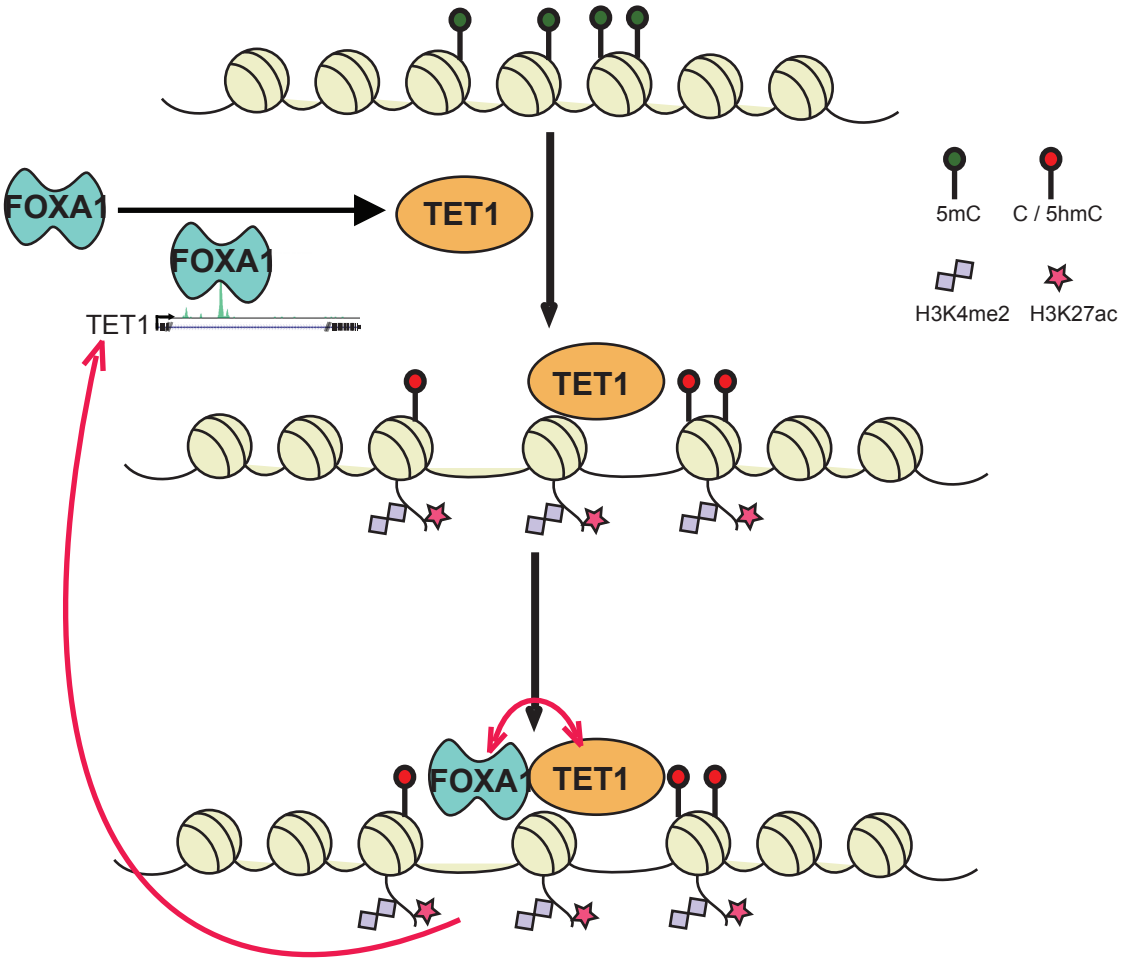
Yang et al. Figure 5



Yang et al. Figure 6



Yang et al. Figure 7



Supplemental Information

FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function

Yeqing Angela Yang[1,*], Jonathan C. Zhao[1,*], Ka-wing Fong[1], Jung Kim[1], Shangze Li[1], Chunxiao Song[2], Bing Song[1], Bin Zheng[1], Chuan He[2,3], Jindan Yu[1,4]

Supplemental Figures:

Supplemental Figure 1. Epigenetic signatures at FOXA1 binding sites.

Supplemental Figure 2. FOXA1 exhibits strong positive correlation with TET1 gene.

Supplemental Figure 3. FOXA1 binds to TET1 enhancer and promoter in C4-2B cells.

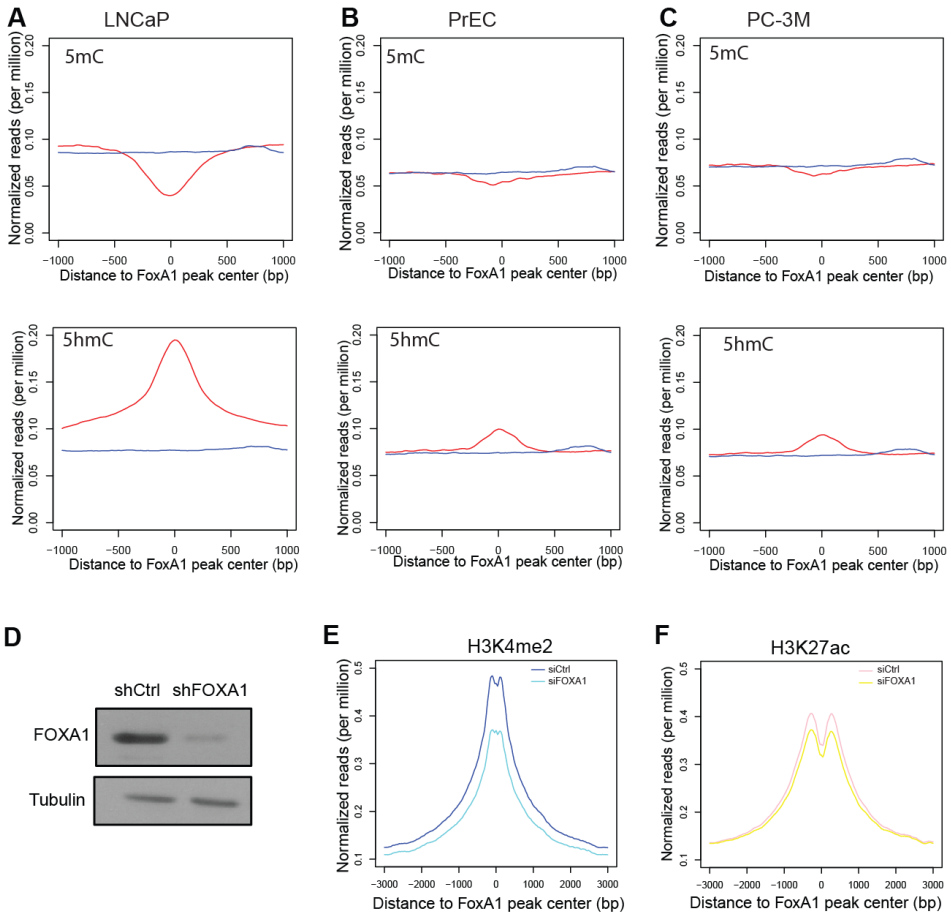
Supplemental Figure 4. DNA-independent interaction between FOXA1 and TET1 proteins.

Supplemental Figure 5. TET1 knockdown reduces 5hmC production.

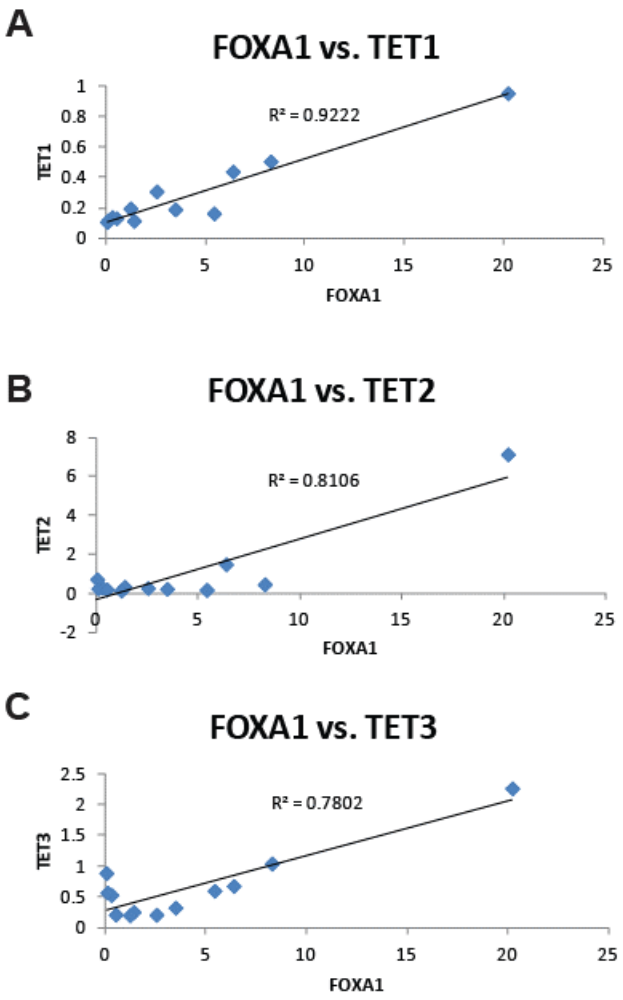
Supplemental Figure 6. TET1 depletion alters the epigenetic signature at FOXA1 binding sites.

Supplemental Methods

Supplemental Table S1: Oligonucleotides that were utilized in this study.



Supplemental Figure 1. Epigenetic signatures at FOXA1 binding sites.
(A-C) Intensity plots showing 5mC and 5hmC enrichment around FOXA1 binding sites (±1 kb) in LNCaP (A), PrEC (B), and PC-3M cells (C). 5mC and 5hmC chemical labeling, or TAmC and hMe-Seal were performed using genomic DNA extracted from LNCaP, PrEC and PC-3M cell lines. Enriched DNA was made into libraries and subjected to deep sequencing. The read intensities of TAmC- and hMe-Seal-seq in different cell lines were evaluated relative to FOXA1 binding sites in LNCaP cells.
(D) Confirmation of FOXA1 knockdown by western blot. LNCaP cells were infected with shCtrl or shFOXA1 lentivirus, and protein lysates were subjected to western analysis using anti-FOXA1 and anti-Tubulin.
(E, F) Intensity plots of H3K4me2 (E) and H3K27ac (F) enrichment around FOXA1 binding sites in control and FOXA1-depleted LNCaP cells. ChIP-seq results for active enhancer marks H3K4me2 and H3K27ac were obtained from publicly available datasets (GSE27823). Enrichment of both histone marks around FOXA1-occupied sites (±3 kb) is shown for control and FOXA1 knockdown LNCaP cells.

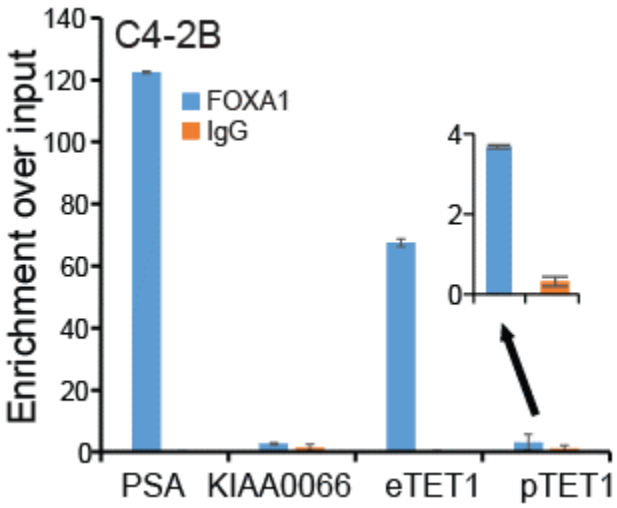


Supplemental Figure 2. FOXA1 exhibits strong positive correlation with TET1 gene.

(A) Scatter plot of FOXA1 and TET1 transcript level, measured by qRT-PCR, in 12 prostate cell lines. Both FOXA1 and TET1 genes were normalized to GAPDH.

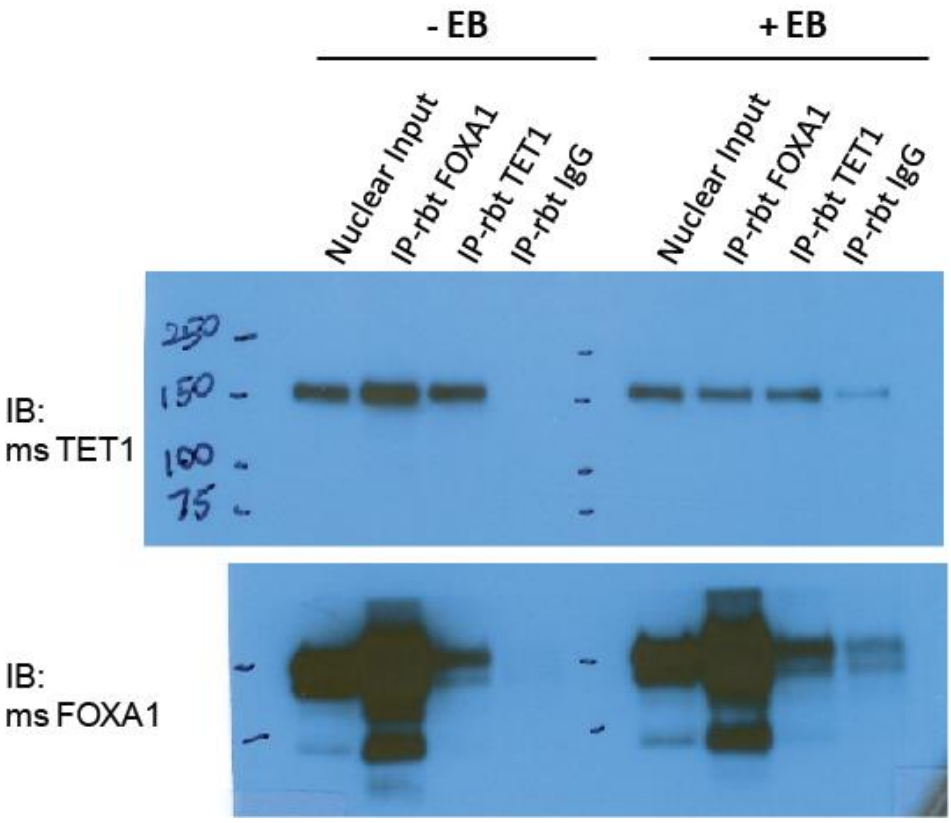
(B) Scatter plot of FOXA1 and TET2 transcript level, measured by qRT-PCR, in 12 prostate cell lines. Both FOXA1 and TET2 genes were normalized to GAPDH.

(C) Scatter plot of FOXA1 and TET3 transcript level, measured by qRT-PCR, in 12 prostate cell lines. Both FOXA1 and TET3 genes were normalized to GAPDH.



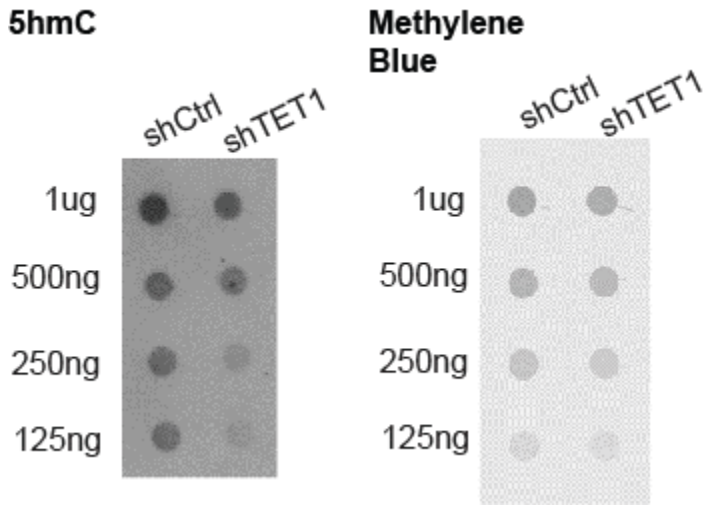
Supplemental Figure 3. FOXA1 binds to TET1 enhancer and promoter in C4-2B cells.

ChIP experiments were done in FOXA1-containing C4-2B cells, using anti-FOXA1 and anti-IgG antibodies. Primers for both TET1 enhancer (eTET1) and promoter (pTET1) regions were used for qPCR analysis. PSA is a positive control gene while KIAA0066 is a negative control gene.



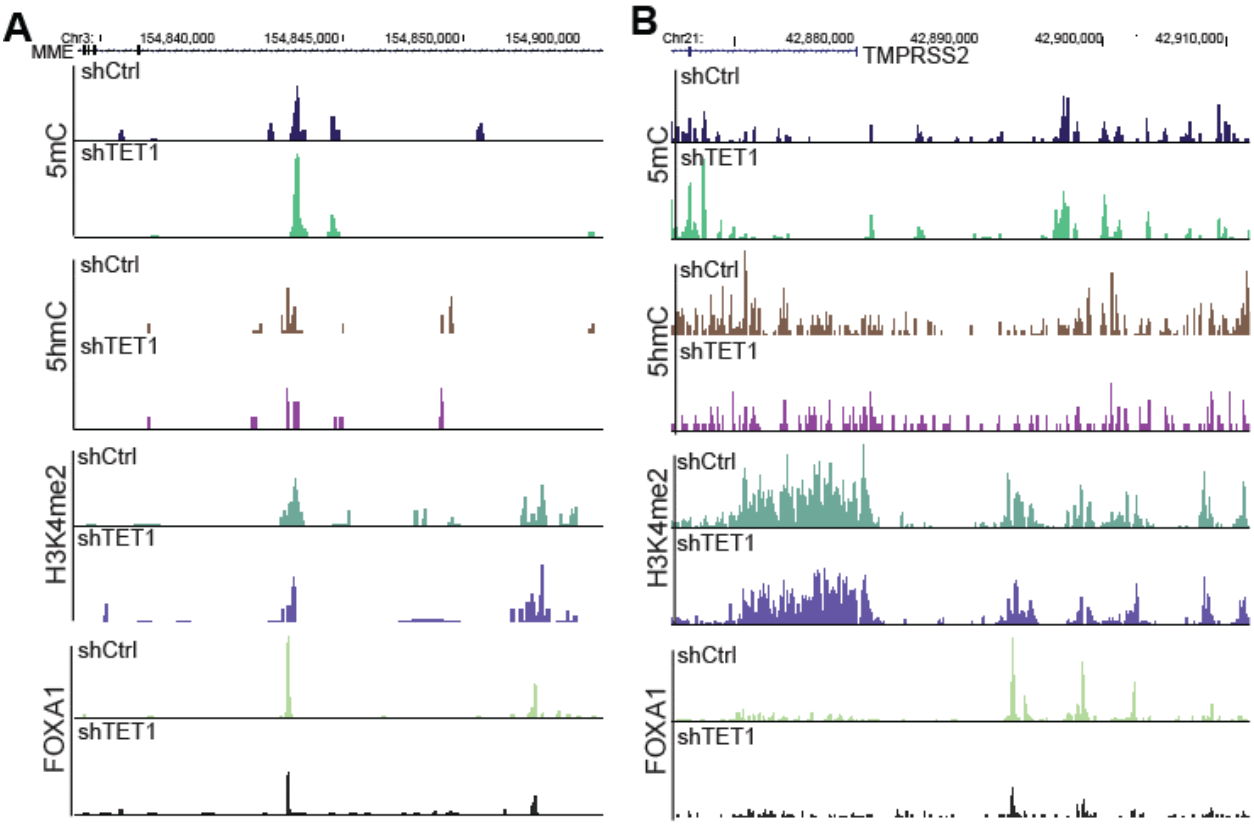
Supplemental Figure 4. DNA-independent interaction between FOXA1 and TET1 proteins.

LNCaP endogenous Co-IP was performed in absence and presence of Ethidium bromide (EB). Nuclear proteins of LNCaP cells were used for IP with anti-FOXA1, anti-TET1 and anti-rabbit IgG antibodies. Ethidium bromide treatment (50ug/ml for 30min) was done to abrogate any potential DNA-mediated protein-protein interaction. Western blot analysis was subsequently used to look at whether FOXA1-TET1 protein interaction is dependent on DNA association.



Supplemental Figure 5. TET1 knockdown reduces 5hmC production.

Genomic DNA extracted from LNCaP control and shTET1 cells was serially diluted and used for dot blot with anti-5hmC antibody. Methylene blue staining was included as loading control.



Supplemental Figure 6. TET1 depletion alters the epigenetic signature at FOXA1 binding sites.

(A, B) Genome browser views of MeDIP- , hMe-Seal- , H3K4me2 and FOXA1 ChIP-seq tracks in representative regions of FOXA1 occupancy, MME enhancer (A) and TMPRSS2 enhancer (B) in control and TET1 knockdown LNCaP cells.

Supplemental Methods:

Nuclear protein extraction

For cell pellets with a packed cell volume of 100ul, 1ml of Lysis Buffer A (10mM HEPES pH 7.9, 10mM KCl, 1.5mM $MgCl_2$, 10% Glycerol, 1mM EDTA pH 8.0) with freshly added protease inhibitors (10% PMSF, cOmplete Protease Inhibitor Cocktail from Roche, 1mM $NaVO_4$, 10mM NaF), was used to resuspend. After 10min of incubation on ice, 0.5% TritonX-100 was added, and sample was vortexed for 15sec to lyse the cells. The homogenate was then centrifuged at 2,000g for 2min at 4^oC. The supernatant contained cytoplasmic proteins and was discarded. The nucleus-containing pellet was resuspended with 300ul Lysis Buffer B (20mM Tris-HCl,pH 7.5, 420mM NaCl, 1.5mM $MgCl_2$, 1% TritonX-100, 10% Glycerol, 1mM EDTA pH 8.0) with freshly added protease inhibitors (10% PMSF, Complete Protease Inhibitor Cocktail from Roche, 1mM $NaVO_4$, 10mM NaF). The sample was incubated on ice for 60min, with periodic vortexing every 10min. The final nuclear lysate was obtained by centrifuging for 10min at 14,000g at 4^oC, after which the supernatant was transferred to a new tube and diluted with Lysis Buffer D (20mM Tris-HCl pH 7.5, 1.5mM $MgCl_2$, 1% TritonX-100, 10% Glycerol, 1mM EDTA pH 8.0), with freshly added protease inhibitors (same as above), to adjust the salt concentration.

Supplemental Table S1: Oligonucleotides that were utilized in this study.

Name	Sequence
GPR137B F	CCCTACTGGGGCACTGTCTA
GPR137B R	TTGCAGGGTACAGCCTCTCT
MME F	TCCTTGAGCTGTGGTGGACT
MME R	CTACGCCCACGGAATCTC
FAT1 F	GGTTCCAAGCAAGACAATCC
FAT1 R	TAGCAGCTGAAGGGTGTGTG
CXADR F	CGCAACCTAGATGCACACAG
CXADR R	AGACAGGGTTTCACCACATTG
NPC1 F	GGATAGGGAAGCTTCTTTCAA
NPC1 R	TTAGGCAGGATGGTCTCGAT
SNAIL F	GGGTTACACCCGTGAACAAG
SNAIL R	CTGGCACCCTTTCATTCTGT
FN1 F	CGCATCTCTTTCCTGTCCAT
FN1 R	GAGGCACCACGAGAAGTGAC
PNLIP F	TGATGTTCCCACAACAATGA
PNLIP R	CATGCACATTGGAAGGTGAG
CNTNAP2 F	GGCAGGATTTCCTCAAAGAC
CNTNAP2 R	GACATCAGCTATCCCCAGGA
TMPRSS2 F	TGGAGCTAGTGCTGCATGTC
TMPRSS2 R	CTGCCTTGCTGTGTGAAAAA
FKBP5 F	GGTTCCTGGGCAGGAGTAAG
FKBP5 R	AACGTGGATCCCACACTCTC
PSA F	GCCTGGATCTGAGAGAGATATCATC
PSA R	ACACCTTTTTTTTTCTGGATTGTTG
TET1 enhancer F	CTCAAGCAATCCTCTTGTCTAGG
TET1 enhancer R	TACACACTGAGTTCAGAGCAAGC
TET1 promoter F	GAACACAGCCCTCATCTGGT
TET1 promoter R	AGAAGGTGCCAGGTCAGAGA
KIAA0066 F	CTAGGAGGGTGGAGGTAGGG
KIAA0066 R	GCCCCAAACAGGAGTAATGA